Psychometrika

A JOURNAL DEVOTED TO THE DEVELOPMENT OF PSYCHOLOGY AS A QUANTITATIVE RATIONAL SCIENCE

THE PSYCHOMETRIC SOCIETY - ORGANIZED IN 1935

VOLUME 21
NUMBER 2
SEPTEMBER
1 9 5 6

PSYCHOMETRIKA, the official journal of the Psychometric Society, is devoted to the development of psychology as a quantitative rational science. Issued four times a year, on March 15, June 15, September 15, and December 15.

SEPTEMBER, 1956, VOLUME 21, NUMBER 3

Published by the Psychometric Society at 1407 Sherwood Avenue, Richmond 5, Virginia. Second-class mail privileges authorized at Richmond, Virginia. Editorial Office, Department of Psychology, The University of North Carolina, Chapel Hill, North Carolina.

Subscription Price: The regular subscription rate is $14.00 per volume. The subscriber receives each issue as it comes out, and, upon request, a second complete set for binding at the end of the year. All annual subscriptions start with the March issue and cover the calendar year. All back issues but two are available. Back issues are $14.00 per volume (one set only) or $3.50 per issue, with a 20 per cent discount to Psychometric Society Members. Members of the Psychometric Society pay annual dues of $7.00, of which $6.30 is in payment of a subscription to *Psychometrika*. Student members of the Psychometric Society pay annual dues of $4.00, of which $3.60 is in payment for the journal.

Application for membership and student membership in the Psychometric Society, together with a check for dues for the calendar year in which application is made, should be sent to

LEONARD KOGAN
105 E. 22nd St., New York 10, N. Y.

Payments: All bills and orders are payable in advance.

Checks covering membership dues should be made payable to the *Psychometric Society*.

Checks covering regular subscriptions to *Psychometrika* (for non-members of the Psychometric Society) and back issue orders should be made payable to the *Psychometric Corporation*. All checks, notices of change of address, and business communications should be sent to

DAVID R. SAUNDERS, Treasurer, Psychometric Society and Psychometric Corporation
Educational Testing Service
P.O. Box 592
Princeton, New Jersey

Articles on the following subjects are published in *Psychometrika:*

(1) the development of quantitative rationale for the solution of psychological problems;
(2) general theoretical articles on quantitative methodology in the social and biological sciences;
(3) new mathematical and statistical techniques for the evaluation of psychological data;
(4) aids in the application of statistical techniques, such as nomographs, tables, work-sheet layouts, forms, and apparatus;
(5) critiques or reviews of significant studies involving the use of quantitative techniques.

The emphasis is to be placed on articles of type (1), in so far as articles of this type are available.

(Continued on the back inside cover page)

Psychometrika

CONTENTS

VOLUME TWENTY-ONE SEPTEMBER 1956 NUMBER 3

PSYCHOMETRIKA—VOL. 21, NO. 3
SEPTEMBER, 1956

THE RELATION BETWEEN INFORMATION AND VARIANCE ANALYSES*

W. R. GARNER

THE JOHNS HOPKINS UNIVERSITY

AND

WILLIAM J. MCGILL

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Analysis of variance and uncertainty analysis are analogous techniques for partitioning variability. In both analyses negative interaction terms due to negative covariance terms that appear when non-orthogonal predictor variables are allowed may occur, Uncertainties can be estimated directly from variances if the form of distribution is assumed. The decision as to which of the techniques to use depends partly on the properties of the criterion variable. Only uncertainty analysis may be used with a non-metric criterion. Since uncertainties are dimensionless (using no metric), however, uncertainty analysis has a generality which may make it useful even when variances can be computed.

I. *Introduction*

Shannon (5) has defined amount of information by the formula

$$H(y) = -\sum_{k=1}^{r} p(k) \log_2 p(k), \tag{1}$$

where y has r discrete values, and $p(k)$ is a probability distribution defined over y. In communication theory, y is considered a source of signals, and the measure H represents the average number of binary digits required to code or store one of the signals. A broader interpretation, however, makes H a parameter which measures the non-metric variability of any probability distribution. H has a value of zero when the probability is concentrated in a single category and is maximum when the probability is uniformly distributed over all categories.

Psychologists have been attracted by the non-metric character of this measure and the obvious application to situations where variances cannot be computed. Since this use of the measure is concerned only with its statistical properties and not with its interpretation in communication theory, we

*The work of the senior author was supported by Contract N5ori-166, Task Order 1, between the U.S. Office of Naval Research and The Johns Hopkins University. This is Report No. 166-I-192, Project Designation No. NR 145-089, under that contract.

shall use the more general term uncertainty, U, to refer to the measure. We shall show that uncertainty has many of the properties of variance and can be partitioned into components as variance can.

II. *The Analysis Problem*

The relations discussed apply when a criterion is predicted from one or more predictors. The development will be presented for the three-variable case, where the problem is to determine to what extent values of the criterion variable can be predicted from two predictor variables.

Our notation is as follows: The criterion variable y can assume any value y_k. The two predictor variables w and x can assume values w_i or x_j. We assume that all three variables are categorized in order that the formulas for uncertainty and variance analysis may have equivalent notations. This assumption does not limit any of the principles demonstrated.

In the three-dimensional matrix, n_{ijk} refers to the number of cases in a single cell; $n_{ij.}$ refers to the total number of cases having the ith value of w and the jth value of x; and $n_{i..}$ refers to the total number of cases having the ith value of w. Similar subscripts indicate other combinations of the three variables; n with no subscript indicates the total number of cases in the matrix. In analysis of variance formulas, $\bar{y}$ indicates a mean value, and the subscript notation just illustrated is used for mean values of the subclassifications.

III. *The Nature of Uncertainty Analysis*

Analysis of variance can be considered as two separate processes. First, the variance of the criterion variable is partitioned into its several identifiable components—components which add up to the total variance. This process is a simple descriptive one; there are no probability assumptions involved in its use. One describes the components of a total variance, making no assumptions about the distributions from which the data are drawn. The second process, which is not a necessary consequence of the first, involves using these partitioned components to obtain estimates of population variances and to make inferences about the parent population. For this process, the actual data provide sample estimates of population distributions; here assumptions about the population distributions become critical.

Uncertainty analysis likewise has both processes. The first process is purely descriptive: it is intended to allow the partitioning of the uncertainty of the criterion variable $U(y)$ into components. Since this process is entirely descriptive, there are no underlying probability assumptions. All that is required for its use is that a data matrix of the type described above is available. The primary purpose of this paper is to demonstrate the nature of uncertainty partitioning and to compare it to variance partitioning. This process is illustrated and explained in Table 1. The results of uncertainty

partitioning specify sources and magnitudes of variabilities as well as amount of categorical discrimination available. These uses are explained more fully by Garner and Hake (1) and by McGill (3).

The n_{ijk} can be considered as sample estimates of $p(i, j, k)$; we can use these sample estimates to test various hypotheses about the parent distribution. For example, suppose we wish to test the hypothesis that both predictors are independent of the criterion, i.e.,

$$p(i, j, k) = p(i, j)p(k). \tag{2}$$

It can be shown (3, 4) by using the likelihood ratio, that when hypothesis (2) is true, [1.3863 $nU(y: w, x)$] is distributed approximately as chi square. Independent tests can be constructed in the same way for each of the predictors separately as well as for the interaction between predictors. The approximation to chi square is of the same order as the familiar chi-square contingency test so that, in effect, uncertainty analysis is analysis of contingency chi square. Miller and Madow (4) discuss this aspect of uncertainty analysis more thoroughly.

IV. *The Orthogonal Case*

Usually in analysis of variance and in uncertainty analysis, the experimenter tries to set up orthogonal predictions. Orthogonality is defined as zero association between the predictors. This requirement is met when the cell frequencies in the matrix of the $n_{ij.}$ can be predicted correctly from the row and column marginal frequencies, i.e., when

$$n_{ij.} = \frac{n_{i..}n_{.j.}}{n}. \tag{3}$$

Uncertainty Analysis

The partitioning of $U(y)$ in uncertainty analysis is illustrated by

$$U(y) = U(y: w, x) + U_{wx}(y), \tag{4}$$

where the uncertainty measures have the definitions given in Table 1. The second term on the right-hand side of (4) is the error uncertainty, i.e., the amount of uncertainty in the criterion y remaining after the predictable uncertainty has been eliminated. The first term on the right-hand side of (4) is the predictable uncertainty; it in turn can be partitioned into components

$$U(y: w, x) = U(y: w) + U(y: x) + U(y: \overline{wx}). \tag{5}$$

These terms are also defined in Table 1. A feature of uncertainty analysis is the interaction term $U(y: \bar{w}\bar{x})$. This is the uncertainty in y predictable from unique combinations of w and x.

Equation (5) describes a process that is identical in form with the

TABLE 1

Symbols, Formulas, and Definitions Used in Three-variable Uncertainty Analysis

The criterion variable, y, is assumed to be non-metric. The predictor variables, w and x, are categorized and may or may not be metric variables.

	Symbol	Formula	Definition
(1)	$U(y)$	$-\sum_{k}\left[\frac{n_{..k}}{n}\right]\log_2\left[\frac{n_{..k}}{n}\right]$	*Total uncertainty:* The amount of uncertainty in the criterion variable, y.
(2)	$U_w(y)$	$-\sum_{i}\left[\frac{n_{i..}}{n}\right]\sum_{k}\left[\frac{n_{i.k}}{n_{i..}}\right]\log_2\left[\frac{n_{i.k}}{n_{i..}}\right]$	*Conditional uncertainty:* The amount of uncertainty in y when one predictor variable, w, is held constant.
(3)	$U_{wx}(y)$	$-\sum_{i,j}\left[\frac{n_{ij.}}{n}\right]\sum_{k}\left[\frac{n_{ijk}}{n_{ij.}}\right]\log_2\left[\frac{n_{ijk}}{n_{ij.}}\right]$	*Error uncertainty:* The amount of uncertainty remaining when both predictor variables, w and x, are held constant.
(4)	$U(y:w)$	$U(y)-U_w(y)$	*Contingent uncertainty:* The uncertainty in y due to the predictor variable, w.
(5)	$U_w(y:x)$	$U_w(y)-U_{wx}(y)$	*Partial contingent uncertainty:* The uncertainty in y due to the predictor variable, x, when the predictor variable, w, is held constant.
(6)	$U(y:w,x)$	$U(y)-U_{wx}(y)$	*Multiple contingent uncertainty:* The uncertainty in y due to the joint influence of the predictor variables, w and x.
(7)	$U(y:\overline{wx})$	$-U(y)+U_w(y)+U_x(y)-U_{wx}(y)$, *or* $U_w(y:x)-U(y:x)$	*Interaction uncertainty:* The uncertainty in y due to unique combinations of the predictor variables, w and x.

TABLE 2

Symbols, Formulas, and Definitions Used in Three-variable Analysis of Variance

The criterion variable, y, is assumed to be metric. The predictor variables, w and x, are categorized but not necessarily metric.

	Symbol	Formula	Definition
(1)	$V(y)$	$\sum_k \left[\frac{n_{..k}}{n}\right] (y_k - \bar{y})^2$	*Total variance:* The variance of the criterion variable, y.
(2)	$V_w(y)$	$\sum_i \left[\frac{n_{i..}}{n}\right] \sum_k \left[\frac{n_{i \cdot k}}{n_{i..}}\right] (y_k - \bar{y}_{i..})^2$	*Conditional variance:* The variance of y when one predictor variable, w, is held constant.
(3)	$V_{wx}(y)$	$\sum_{i,j} \left[\frac{n_{ij.}}{n}\right] \sum_k \left[\frac{n_{ijk}}{n_{ij.}}\right] (y_k - \bar{y}_{ij.})^2$	*Error variance:* The variance remaining in y when both predictor variables, w and x, are held constant.
(4)	$V(y:w)$	$V(y) - V_w(y)$	*Main effect:* The variance of y due to the predictor variable, w.
(5)	$V_w(y:x)$	$V_w(y) - V_{wx}(y)$	*Partial main effect:* The variance of y due to the predictor variable x, when the predictor variable, w, is held constant.
(6)	$V(y:w,x)$	$V(y) - V_{wx}(y)$	*Total predictable variance:* The variance of y due to the joint influence of the predictor variables, w and x.
(7)	$V(y:\overline{wx})$	$-V(y) + V_w(y) + V_x(y) - V_{wx}(y)$, or $V_w(y:x) - V(y:x)$	*Interaction variance:* The variance of y due to unique combinations of the predictor variables, w and x.

partitioning of variance in analysis of variance; in the orthogonal case the interaction uncertainty can be interpreted by analogy with interaction variance. This is true despite the fact that interaction uncertainties are sometimes negative (**3**). This problem will be discussed in detail in Section V.

Analysis of Variance

Uncertainty analysis is generally appropriate when the criterion variable y is a categorical variable, i.e., one allowing only nominal scale values (cf. **6**). The predictor variables may be categorical, or they may be metric variables which are categorized for purposes of analysis. If the criterion is a true metric variable, i.e., one having at least the properties of an interval scale, we can compute variances and perform analysis of variance. The predictor variables must be categorized in any simple form of the analysis of variance.

Equations describing analysis of variance are essentially identical to those of uncertainty analysis. The defining equations are given in Table 2; except for the fact that variances are computed from squared deviations, whereas uncertainties are computed from log-probabilities, the equations are identical to those in Table 1. The partition of the variance of the criterion can be written:

$$V(y) = V(y: w, x) + V_{wx}(y). \tag{6}$$

Again the two parts on the right-hand side of the equation are the predictable and the error components of the total variance. The predictable variance can be broken down as before:

$$V(y: w, x) = V(y: w) + V(y: x) + V(y: \overline{wx}). \tag{7}$$

The terms in (7) are explained in detail in Table 2.

Normally the analysis of variance in (7) is called double classification; the variances are generally identified in terms of the two predictors. This shorthand procedure is convenient for most purposes. However, it obscures the fact that the data array is three-dimensional. The analysis is identical to the one treated in uncertainty analysis in every respect, except that in the analysis of variance the criterion variable has a metric, whereas it does not in uncertainty analysis.

V. *The Non-Orthogonal Case*

In Section IV it was mentioned that the interaction term in uncertainty analysis can assume negative values under certain conditions. It is equally true that the interaction term in analysis of variance can be negative, if it is defined as in Table 2. The negative interaction term is due to non-orthogonality and can be thought of as due to a negative covariance term that may attenuate or exceed the positive interaction effect.

Uncertainty Analysis

It is not difficult to show that the interaction uncertainty in (5) can be written

$$U(y: \overline{wx}) = U_y(w: x) - U(w: x). \tag{8}$$

This form of the interaction term shows at once that interaction cannot be negative with orthogonal predictors since orthogonality requires that $U(w: x) = 0$.

In the non-orthogonal case, however, $U(w: x)$ will be greater than zero. With certain combinations of cell frequencies, the contingent uncertainty between x and w can be larger than the partial contingent uncertainty—resulting in negative interaction. A simple illustration of this principle is provided when each value of w is paired uniquely with each value of x. Now $U(w: x)$ is as large as it can be. Furthermore, $U_y(w: x)$ cannot be greater than $U(w: x)$ since $U(w: x)$ is the maximum contingent uncertainty that can be obtained from a contingency table involving w and x. Equation (8) shows that the interaction will never be greater than zero. An identical result is obtained in the variance analysis when the predictors are completely confounded.

Analysis of Variance

It is usually assumed that the components of the total variance in analysis of variance must be positive. This is true only in the orthogonal case; if an analysis of variance is carried out with a non-orthogonal experimental design, using the equations given in Table 2, negative interaction terms can occur.

To show how this happens, we now analyze the components of the interaction variance for the general case. The equation is

$$V(y: \overline{wx}) = \frac{1}{n} \sum_{i,j} n_{ij.}(\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y})^2 - \frac{2}{n} \sum_{i,j} \left(n_{ij.} - \frac{n_{i..} n_{.j.}}{n} \right) (\bar{y}_{i..} - \bar{y})(\bar{y}_{.j.} - \bar{y}). \tag{9}$$

It can be seen that the interaction variance is composed of two parts: the first part is essentially the interaction variance in the orthogonal case; the second part is a negative covariance term. This term must be zero in the orthogonal case [see equation (3)], but in the non-orthogonal case it cannot be ignored. The redundancy introduced by non-orthogonality is illustrated clearly in multiple regression. No interaction term is permitted, but a correction for non-orthogonality must be introduced whenever the predictor variables are correlated (cf. **2**).

VI. *Effects of Non-Orthogonality*

Our discussion of non-orthogonality shows that it is best to design experiments with orthogonal predictor variables. The analysis is simplified, and the uninterpretable interaction components are eliminated.

Clearly the covariance in (9) is not just part of the interaction variance. In fact, when predictors are non-orthogonal, the concept of interaction is almost meaningless. For example, consider an analysis of variance in which the predictors w and x are completely confounded. The two main-effect variances and the interaction will all be identical. The covariance term in (9) must be large enough to cancel out two of these variances, but we do not know which two of the variances should be cancelled out. In a sense, the covariance term is a correction factor which must be applied to the entire set of variances. Thus, a covariance term (whether or not it is large enough to produce a negative interaction) renders an exact interpretation of the component variances impossible.

The multiple contingent uncertainty or the total predictable variance can be computed directly as shown in the defining equations in Tables 1 and 2. The negative covariance term is included; there is no over-estimation of the total predictable variance or uncertainty. However, the interpretation of results should be made only in terms of combinations of the two predictors —no valid statements can be made about them independently.

Sometimes it is impossible to obtain orthogonal predictor variables, particularly when there are more than two. In time series successive events are usually not orthogonally related because no independent control of these events is possible. If the time series has serial dependencies, preceding events cannot be orthogonal. Consequently, the total predictability of events in a time series cannot in general be computed by adding up the separate predictabilities obtained from preceding events displayed by one or more units in the time series.

VII. *Estimation of Uncertainties from Variances*

It is clear that uncertainty analysis and analysis of variance are analogous analytic techniques. In fact, variances may be used to estimate uncertainties if we assume that y is normally distributed.

Shannon (5) has shown that the uncertainty of a normal distribution can be specified as

$$\text{est } U(y) = \tfrac{1}{2} \log_2 2\pi e V(y) - \log_2 m, \tag{10}$$

where est $U(y)$ is the estimated total uncertainty of the criterion variable on the assumption of a normal distribution of values of y_k, and where m is the width of the category interval on the y continuum.

We can write similar equations for any of the variances obtained in

analysis of variance. For example,

$$\text{est } U_{wx}(y) = \tfrac{1}{2} \log_2 2\pi e V_{wx}(y) - \log_2 m \tag{11}$$

is the error uncertainty estimated from error variance. From definition (6) in Table 1, and from equations (10) and (11), we can write

$$\text{est } U(y: w, x) = \tfrac{1}{2} \log_2 [V(y)/V_{wx}(y)]. \tag{12}$$

Thus, it is relatively simple to estimate the multiple contingent uncertainty from the appropriate variances. The expression on the right-hand side of this equation is reminiscent of the multiple correlation ratio (η). We can, in fact, write

$$\text{est } U(y: w, x) = -\tfrac{1}{2} \log_2 [1 - \eta^2(y: w, x)]. \tag{12-A}$$

Estimated uncertainties have the properties of additivity observed in computed uncertainties. Consequently, the expression on the right-hand side of (12) can be partitioned into three components, each of which is based on its equivalent variances as follows:

$$\text{est } U(y: w) = \tfrac{1}{2} \log_2 [V(y)/V_w(y)], \tag{13}$$

$$\text{est } U(y: x) = \tfrac{1}{2} \log_2 [V(y)/V_x(y)], \tag{14}$$

$$\text{est } U(y: \overline{wx}) = \tfrac{1}{2} \log_2 \{[V_w(y) \cdot V_x(y)]/[V(y) \cdot V_{wx}(y)]\}. \tag{15}$$

These estimating equations point out some of the differences between uncertainty and variance. If (15) is used to estimate the interaction uncertainty when the interaction variance is zero, cases can be found in which the estimated interaction uncertainty (and the computed interaction uncertainty) will not be zero. Converse cases (i.e., zero uncertainty interactions with finite variance interactions) can also be found. These apparent contradictions are due to the fact that variances and uncertainties, while analogous, do not measure exactly the same characteristics of probability distributions. Uncertainty analysis depends on the number of categories occupied by a distribution. Variance analysis depends on the weights or values attached to these categories.

VIII. *Application of the Measures*

We have now shown that uncertainty analysis and analysis of variance are equivalent in many respects; the question naturally arises as to when one should be used in preference to the other. This decision depends on the properties of the data and the assumptions the experimenter is willing to make. If the criterion variable y has only the properties of a nominal or ordinal scale, then only uncertainty analysis is permissible. Uncertainty analysis has the greater generality and requires no assumptions about metric properties of the criterion.

On the other hand, uncertainty analysis does not give any information about the metric if it exists. If the criterion variable is metric with at least the properties of an interval scale, then analysis of variance must be used to retain information about the metric. The variance measure in retaining the metric sacrifices generality since the variances obtained from one experiment are not directly comparable to those obtained from another. Thus, the fact that the uncertainty measure is dimensionless gives it a generality which allows direct comparison of experimental results which differ in their metric.

To summarize, the measures are similar in many respects, but they are not identical. The uncertainty measure has greater generality and the advantages of generality. The variance measure is more specific but retains information about the metric. The decision as to which to use depends not only upon the properties of the criterion variable but also upon the gain expected from being more sensitive instead of more general. In many applications it is reasonable to use both measures and compare them.

REFERENCES

1. Garner, W. R. and Hake, H. W. The amount of information in absolute judgments. *Psychol. Rev.*, 1951, **58**, 446-459.
2. Garner, W. R. and McGill, W. J. Relation between uncertainty, variance, and correlation analyses. Mimeographed report, Department of Psychology, The Johns Hopkins University, Baltimore, Maryland.
3. McGill, W. J. Multivariate information transmission. *Psychometrika*, 1954, **19**, 97-116.
4. Miller, G. A. and Madow, W. G. On the maximum likelihood estimate of the Shannon-Wiener measure of information. Report AFCRC-TR-54-75. Operational Applications Laboratory, ARDC, August 1954.
5. Shannon, C. E. A mathematical theory of communication. *Bell syst. tech. J.*, 1948, **27**, 379-423, 623-656.
6. Stevens, S. S. Mathematics, measurement, and psychophysics. In S. S. Stevens (Ed.), Handbook of experimental psychology. New York: Wiley, 1951, pp. 1-49.

Manuscript received 10/9/54

Revised manuscript received 9/23/55

PSYCHOMETRIKA—VOL. 21, NO. 3
SEPTEMBER, 1956

MEASUREMENT OF SUBJECTIVE VALUES

HAROLD GULLIKSEN

PRINCETON UNIVERSITY
AND
EDUCATIONAL TESTING SERVICE

Four different value laws are developed and tested by using them to predict the scale values of composite stimuli from the scale values of their components. These four laws are: an additive law, a square-root law, a logarithmic, and a negative exponential law. They are tried out on a set of food preferences by means of Pearson's Method of False Position. The negative exponential law of diminishing returns gave the best fit to the data but was not markedly better than any of the other laws.

The purpose of this study is to show that laws relating subjective value to amount of commodity may be studied by an extension of the usual psychophysical scaling methods. Four different value laws will be developed and tested with a set of experimental data on food preferences.

The psychophysical scaling procedures, such as paired comparisons, for example, may be used to distinguish between various types of laws expressing value increase as a function of increase in amount of the commodity. These procedures are applicable even when no physical measurement of the amount of the commodity can be made, and when the scaling procedure necessitates measuring from an arbitrary origin.

Testing each of these laws of value increase also involves a corresponding determination of an origin, or point of zero value. Various psychophysical methods of determining a zero point have been presented in the literature. An additive law of value increase has been used for this purpose (cf. **1**, **4**, and **7**). It will be shown here that other laws of value increase may also be used to determine an origin or point of zero value.

The data necessary to test these formulations are obtained by using a preference schedule like that illustrated in Table 1. The subject is asked the usual paired comparison question, "Which do you prefer, i or j?" However, in addition to the single stimuli i and j *composite* stimuli of the form (i and j) or (g and h) are used. The subject is asked questions of the form, "Do you prefer (i and j) or g?" as well as, "Do you prefer (i and j) or (g and h)?" (**15**).

Do these different stimuli, designated by i, j, g, h, (i and j), (g and h), etc., behave as if they are different amounts of some commodity x whose subjective value v is given by some function, say $v = f(x)$? For any given function, the experimental device of utilizing the *composite* stimuli of the

type (i and j) may be indicated by writing

$$v_i = f(x_i); \quad v_j = f(x_j); \quad \text{and} \quad v_{ij} = f(x_i + x_j). \tag{A}$$

For example, if the items i, j, and (i and j) were bundles of dollars, such that the number of dollars in each bundle were known to the subject, but not to the experimenter, then the experimenter could raise the question asked here. Do these different bundles and their combinations behave as if they were different packages containing different numbers of dollars x, which were known to the subject and were related to the subjective value v by some function $v = f(x)$?

It is of interest to note that since

$$x_i = f^{-1}(v_i); \quad x_j = f^{-1}(v_j); \quad \text{and} \quad x_i + x_j = f^{-1}(v_{ij}), \tag{B}$$

we have

$$f^{-1}(v_{ij}) = f^{-1}(v_i) + f^{-1}(v_j). \tag{C}$$

From the viewpoint of psychological scaling, we do not obtain v's, but instead obtain s's which differ from the v's by a constant, so we may write

$$f^{-1}(s_{ij} - c) = f^{-1}(s_i - c) + f^{-1}(s_j - c). \tag{D}$$

In other words, if one can find any function, designated f^{-1}, such that an additive relationship holds for the proper scale values as indicated then one possible solution for the relationship between v and x is the inverse f of this function f^{-1}. From this point of view the commodity amounts are simply defined by $x_i = f^{-1}(v_i)$.

If another function g is found such that

$$g^{-1}(v_{ij}) = g^{-1}(v_i) + g^{-1}(v_j), \tag{E}$$

we may regard it as defining another commodity amount $y_i = g^{-1}(v_i)$. It is of interest to ask about the relationship between x_i and y_i or between g^{-1} and f^{-1}. Substituting (A) in (E),

$$g^{-1}f(x_i + x_j) = g^{-1}f(x_i) + g^{-1}f(x_j). \tag{F}$$

A theorem in functional equations states that if

$$F(X + Y) = F(X) + F(Y), \tag{G}$$

then

$$F(X) = aX, \tag{H}$$

where a is a constant coefficient. This theorem was proved for continuous functions by Cauchy (2). The discontinuous solution for (G) is discussed by Hamel (6) and Sierpiński (11). Applying solution (H) to (F) we have

$$g^{-1}f(x_i) = a(x_i).$$

Utilizing (B) gives

$$g^{-1}(v_i) = af^{-1}(v_i) \quad \text{or} \quad y = ax.$$

In other words, the "different" value laws found in this way would differ only in the unit of measurement used for the commodity. It should be noted that this statement holds only if equations (B) and (E) hold without error for values of v_i , v_j , and v_{ij} .

In order to test different value laws, the theoretical formulation will need to show the relationship between the values v_i and v_j for the *component* stimuli and the value v_{ij} of the *composite* stimulus. Let us see what different laws of value increase imply with respect to this relationship.

Four Laws of Value Increase

Logarithmic Law

If it is assumed that the rate of change in value v is inversely proportional to the amount of the commodity x, we have the differential equation

$$dv/dx = k/x. \tag{1}$$

Integrating this equation and setting $x = x_0$ when $v = 0$, we have

$$v = k \log (x/x_0). \tag{2}$$

This derivation of a logarithmic law of value increase was given by Thurstone (**13**).

What does this logarithmic value law imply with respect to the relationship between the values of the *component* and the *composite* stimuli? In order to determine this, we note that

$$\begin{aligned} v_{ij} &= k \log [(x_i/x_0) + (x_j/x_0)], \\ v_i &= k \log (x_i/x_0), \end{aligned} \tag{3}$$

and

$$v_j = k \log (x_j/x_0).$$

Eliminating x_i and x_j among these equations enables us to find an expression for the composite value v_{ij} in terms of the values v_i and v_j of the components. Thus, we find that

$$e^{v_{ij}/k} = e^{v_i/k} + e^{v_j/k}. \tag{4}$$

These equations do not require the measurement of amount of commodity x, but require only a set of value measurements v. However, the usual scaling data give neither the v's nor the x's but only scale values, s, which differ from v by a constant. Let C be the scale value for which v equals zero.

By making the substitutions

$$v_i = s_i - C,$$
$$v_j = s_j - C, \tag{5}$$

and

$$v_{ij} = s_{ij} - C,$$

let us determine the type of relationship among the scale values of the component stimuli and the composite stimuli that is implied by the logarithmic law of value increase. Substituting and simplifying gives

$$e^{s_{ij}/k} = e^{s_i/k} + e^{s_j/k}. \tag{6}$$

This equation gives the interrelationships among the experimentally determinable scale values which are implied by the logarithmic law of value increase.

There are some interesting and disconcerting things about this law. It contains the parameter k, which is involved in such a way that it cannot readily be solved for explicitly. Thus, it remains as an annoying trial parameter in any attempts to verify this equation. It is also interesting to note that the equation does not contain C at all. In other words, for the logarithmic law of value increase the additive constant C cannot be determined. *Any* additive constant is consistent with the relationships among the component and composite scale values. Likewise, any value of x_0 is consistent with these relationships, since for this law, x_0 functions as a unit of measurement for x.

Square-Root Law

If it is assumed that the rate of change in value v is inversely proportional to the value level already attained, we may write the differential equation

$$dv/dx = k/2v. \tag{7}$$

Integrating and setting $x = x_0$ when $v = 0$ we have

$$v = \sqrt{k(x - x_0)}. \tag{8}$$

This derivation of a square-root law of value increase was given by Thurstone (**13**).

To determine the implications of this square-root law with respect to the relationship between the values of the *component* and *composite* stimuli, we note that

$$v_i = \sqrt{k(x_i - x_0)}$$
$$v_j = \sqrt{k(x_j - x_0)} \tag{9}$$
$$v_{ij} = \sqrt{k(x_i + x_j - x_0)}.$$

Eliminating x_i and x_j among these equations enables us to find an expression for the composite value v_{ij} in terms of the values v_i and v_j of the components. Thus, we find that

$$v_{ij}^2 = v_i^2 + v_j^2 + kx_0 . \tag{10}$$

In order to determine the relationship among the scale values implied by this equation, we substitute (5) in (10), giving

$$s_{ij}^2 - s_i^2 - s_j^2 = 2C(s_{ij} - s_i - s_j) + C^2 + kx_0 . \tag{11}$$

This equation gives the interrelationship among the experimentally determinable scale values which is implied by the square-root law of value increase. The plot of the quantity on the left side of the equation against the quantity in parentheses enables one to find C from the slope of the line, and also to find kx_0 by subtracting the square of half the slope from the intercept. Separate values for k and x_0 cannot be determined.

Negative Exponential Law

If it is assumed that the rate of change in value v is directly proportional to the difference between the value level already reached and an asymptotic value level A, we have the differential equation

$$dv/dx = k(A - v). \tag{12}$$

Integrating and setting $x = x_0$ when $v = 0$ gives

$$v = A - Ae^{kx_0}e^{-kx}. \tag{13}$$

This is the familiar negative exponential law of diminishing returns used in economics (see, for example, **8** and **12**). It has also been suggested by a number of writers as an equation of the learning curve (see illustrations cited in **3**). If we apply it to the component stimuli "i" and "j" and also to the composite "i and j", we have

$$\begin{aligned} v_i &= A - Ae^{kx_0}e^{-kx_i}, \\ v_j &= A - Ae^{kx_0}e^{-kx_j}, \\ v_{ij} &= A - Ae^{kx_0}e^{-k(x_i+x_j)}. \end{aligned} \tag{14}$$

Eliminating x_i and x_j to find v_{ij} as a function of v_i and v_j gives

$$v_{ij} = A - (1/A)e^{-kx_0}(A - v_i)(A - v_j). \tag{15}$$

If v_i is plotted against v_{ij} for a given value of v_j , the result is a straight line. If this plot is made for each of the values of v_j , the result is a family of straight lines. We also note that if either v_i or v_j is equal to A, then the right-hand term of the equation vanishes, giving $v_{ij} = A$. Thus, the indicated plots constitute a *pencil* of straight lines intersecting in the point (A, A). Again

we may substitute for the v's in terms of the s's (equations 5), obtaining

$$s_{ij} = A + C - (1/A)e^{-kx_0}(A + C - s_i)(A + C - s_j). \tag{16}$$

Again, if s_i is plotted against s_{ij} for a given value of s_j , the result is a straight line. If such a plot is made in turn for each of the possible values of s_j the result is a *pencil* of straight lines intersecting in the point $(A + C, A + C)$. If $x_0 = 0$ this series of straight lines may be used to give the values of A and of C.

Linear Law

If we assume that the rate of change of value is constant, we have the differential equation

$$dv/dx = k. \tag{17}$$

Integrating and setting $x = x_0$ when $v = 0$, we have

$$v = k(x - x_0). \tag{18}$$

If we solve for the interrelationships among the component and composite scale values implied by this equation, we find that

$$s_{ij} = s_i + s_j - C + kx_0 . \tag{19}$$

This relationship has been derived by Thurstone and utilized in an unpublished study (**15**). According to this law, the plot of the scale value of the composite against the sum of the scale values of the components should be linear with unit slope and intercept $kx_0 - C$.

The Food Preference Experiment

Let us now consider the type of data that is used for these value studies. Food preferences were studied.

Pairs of single items are presented such as *Beef* vs. *Pork*, and the subject is asked to indicate which he would choose, i or j. Then pairs of what we shall term *composite* items are presented. The subject is asked to choose between *Beef and Steak* vs. *Tongue and Lamb*. Also the choice is given between single and composite items, such as *Steak* vs. *Pork and Tongue*. Table 1 shows three typical items in the schedule.

The set of 5 component stimuli and 10 composite stimuli (making 15 stimuli in all) were presented in a paired comparison schedule to 92 college students in a psychology class. The directions stressed that for a composite such as *Beef and Steak, each* is an ordinary sized serving, and that if the composite were chosen, the person was to eat *both*, thus having twice as much as if only a single *component* item were chosen. This was done in order to give a better chance for diminishing returns to be manifested in the results. All choices involving a duplicate item were omitted, resulting in a matrix of

TABLE 1

Sample Questionnaire Items	
38.	☐ Roast Rib of Prime Beef
	☐ Roast Loin of Pork
39.	☐ Roast Rib of Prime Beef Sirloin Steak
	☐ Boiled Smoked Beef Tongue Loin Lamb Chop
40.	☐ Sirloin Steak
	☐ Roast Loin of Pork Boiled Smoked Beef Tongue

incomplete data. A least squares procedure was used (**5**) for scaling a paired comparisons matrix of incomplete data.

The scale values obtained range from .000 for tongue and pork, the least preferred item, on up through 1.043 for lamb, to 2.622 for the composite, beef and steak, which was the most preferred item. The complete set of scale values is shown in Table 2.

The fact that these 15 scale values give a good fit (**5**) to the 55 paired comparisons judgments shows that persons can make consistent judgments about preferences for composite stimuli along with single stimuli of the type used here. Thus, it is experimentally feasible to present in a single schedule comparisons of the (i vs. j), (i vs. g and h), and (i and j vs. g and h) types. Any set of concrete objects or even abstract concepts can be dealt with according to this pattern.

In Table 2 we notice that the value of tongue and lamb is higher than tongue alone, pork and lamb is higher than pork alone, beef and lamb is higher than beef alone, and lamb and steak is higher than steak alone. Thus, the value of any item is increased by forming a composite with lamb. The same holds true for beef and steak. Thus, lamb, beef, and steak are all positive values. However, now look at pork. The value of pork and steak is lower than the value of steak alone, of pork and beef lower than beef alone, of pork and lamb lower than lamb alone, and the value of the *composite* tongue and pork is lower than that of either of the components. Thus, from the purely ordinal characteristics of the scale, it seems clear that pork has a negative value. The same is true to an even greater extent for tongue. The zero point is between pork and lamb since pork and tongue are negative and lamb, beef, and steak are positive.

TABLE 2

Food Preference Experiment

Stimuli	Scale Values
Tongue and Pork	.000
Tongue	.137
Tongue and Lamb	.270
Pork	.541
Tongue and Beef	.830
Pork and Lamb	.928
Lamb	1.043
Tongue and Steak	1.088
Pork and Beef	1.448
Beef	1.746
Pork and Steak	1.780
Lamb and Beef	1.993
Steak	2.197
Lamb and Steak	2.324
Beef and Steak	2.622

Generalization to Include Positive and Negative Values

For this particular set of data there is thus clear evidence that some of the component values are positive and others are negative. The *linear* value law extends readily to include both positive and negative values and their various combinations.

The other laws, however, in their previously stated form do not give reasonable results for both positive and negative values. However, it is possible to make an appropriate extension by having four different rules, depending first on whether the components had the same or different signs, and second on whether the composite was positive or negative.

A reasonable interpretation including both negative and positive values and their combinations for the square-root law may be made by assuming $x_0 = 0$ and writing

$$\hat{s}_{ij} = \frac{v_i + v_j}{|v_i + v_j|}\sqrt{\left|v_i^2 + \frac{v_i v_j}{|v_i v_j|}v_j^2\right|} + C, \tag{20}$$

where $v_i = s_i - C$; $v_j = s_j - \mathrm{C}$.

This formulation merely says that v_i^2 and v_j^2 are added if v_i and v_j are of the same sign and subtracted if v_i and v_j are of unlike signs. The square root of the sum is then given the sign of the larger value. If v_i and v_j are both positive, (20) is equivalent to (10) or (11). If v_i and v_j are both negative, (20) is equivalent to a reflection into the quadrant where s_i , s_j , and s_{ij} are each negative. If v_i and v_j are of opposite sign, (20) gives an interpretation that is consistent with the previous cases.

In order to state the logarithmic law for a series of either positive or negative values we have analogously with the square-root law:

$$\hat{s}_{ij} = \frac{v_i + v_j}{| v_i + v_j |} k \log \left| e^{|v_i|/k} + \frac{v_i v_j}{| v_i v_j |} e^{|v_j|/k} \right| + C, \tag{21}$$

where $v_i = s_i - C$; $v_j = s_j - C$.

Again this formula merely states that e is taken to be a positive power for either positive or negative values. The resulting powers are added if v_i and v_j are of the same sign and subtracted if v_i and v_j have different signs. The logarithm is then given the sign of the larger value. In contrast to (6), it is now possible to determine C for (21) since the *combination* of negative and positive values is involved.

The negative exponential law may be restated for the case in which both negative and positive values are involved. Let $x_0 = 0$; thus

$$\frac{A - v}{A} = e^{-kx} \qquad \text{(where } v = s - C\text{)}. \tag{22}$$

The formulas are easier to work with if we define B as the asymptote expressed in terms of the s-scale, just as C is defined:

$$B = A + C.$$

Let us assume that the positive (upper) asymptote B^+ may vary independently of the negative (lower) asymptote B^-. Then let us define

$$R_i = \frac{B^{s_i} - s_i^{s_i}}{| B^{s_i} - C |} = \frac{x_i}{| x_i |} e^{-k|x_i|}. \tag{23}$$

The superscript signs are used to indicate that the positive asymptote B^+ is used if $s_i > C$ (i.e., for s^+); the negative asymptote B^- is used if $s_i < C$ (i.e., for s^-). The R's thus defined are positive quantities if $s_i > C$, and negative if $s_i < C$. Note that as $| x_i |$ increases $| R_i |$ decreases. Then

$$\hat{s}_{ij} = B_i - (B_i - C) a R_i^a R_i , \tag{24}$$

where $a = R_i R_j / | R_i R_j |$ and $| R_j | > | R_i |$.

Equation (24) gives a set of computations expressed entirely in s-scale values by means of which an estimate $\hat{s}_{ij}$ of the scale value of a composite may be computed from the scale values of the components, assuming a negative exponential law of value increase.

Method of False Position

It was also found that the test equations previously developed were not sensitive enough to differentiate clearly between the different value laws. The method adopted was that of using the component stimuli to predict the composite value—selecting the parameters to minimize the sum of the squares of the differences between the observed and predicted scale values for the composite stimuli. For the linear additive rule, the solution is straightforward. For the others no explicit solution could be found, so the fit was made by a successive approximations procedure using Pearson's Method of False Position (**10**). A brief account of the method presented in matrix notation will be given here. This method is a general solution for linear or nonlinear equations presented by Karl Pearson. The problem may be stated as follows:

Given the k-parameter function

$$Y_i = f(m_1 , m_2 , \cdots , m_g , \cdots , m_k , x_i)$$

together with experimental observations of the paired values x_i , y_i $(i = 1, \cdots , n)$, to determine the values of m_g so as to minimize $\sum_{i=1}^{n} (y_i - Y_i)^2$. Only two restrictive conditions are necessary:

(1) Given the value of the independent variable x_i and a set of arbitrary values m_{hg} for each of the parameters, it must be possible to compute (or to obtain by mechanical or other means) a corresponding set of values Y_{hi} for the dependent variable.

(2) The function $Y_{hi} = f(m_{h1} , m_{h2} , \cdots m_{hg} , \cdots m_{hk} , x_i)$ must be continuous and have continuous derivatives for slight changes in the values of the parameters in the neighborhood of the desired solution.

For a one-parameter function the Law of Mean Value may be stated as

$$\frac{dY_{pi}}{dm_{p0}} = \frac{Y_{hi} - Y_{0i}}{m_h - m_0} ,$$

or

$$Y_{hi} - Y_{0i} = \frac{dY_{pi}}{dm_{p0}} (m_h - m_0).$$

For a k-parameter function the general Law of the Mean for functions of several variables (cf. **9**, p. 121) gives

$$Y_{hi} - Y_{0i} = \sum_{g=1}^{k} \left[\frac{\partial Y_{pi}}{\partial m_{pg}} (m_{hg} - m_{0g}) \right],$$

where

- i is an index for the observations of the dependent and independent variables $(i = 1, \cdots , n)$,
- g is an index for the parameters in the function $(g = 1, \cdots , k)$,
- 0 indicates the initial guess for the values of the parameters,

h indicates subsequent guesses for the value of the parameters ($h = 1, \cdots, k$),

p indicates that the value of the partial derivative is taken at a point p on the curve between m_{0g} and m_{hg} ,

m_{hg} indicates the hth estimate for the parameter m_g .

Pearson's derivation is algebraic and quite voluminous. A much briefer statement in terms of matrices has been given by Gale Young in an unpublished note to the writer. This derivation is presented with acknowledgment to Dr. Young. We may put the derivation in matrix terminology as follows: Let

Y be a matrix of k rows and n columns with elements $Y_{hi} - Y_{0i}$ ($h = 1, \cdots, k$; $i = 1, \cdots, n$).

M be a square matrix with elements $m_{hg} - m_{0g}$, where ($g = 1, \cdots, k$; $h = 1, \cdots, k$).

F_p be a matrix of k rows and n columns with elements $\partial Y_{pi}/\partial m_{pg}$.

Then from the Law of Mean Value

$$Y = MF_p \quad \text{or} \quad M^{-1}Y = F_p .$$

Thus, we have a means of eliminating matrix F_p , for which it would be very difficult to find reasonable experimental values. In order for M^{-1} to exist, M must be a square matrix; hence for k parameters there must be $k + 1$ guesses for each parameter. Thus, both g and h must vary from 1 to k. It should also be noted that the partial derivatives in F_p are taken at some suitable point p between m_{0g} and m_{hg} , selected so that the Law of Mean Value holds. Let

y_i ($i = 1, \cdots, n$) designate the set of observed y values,

m_{bg} designate the parameter values which give the best fit,

Y_{bi} designate the corresponding values for the best fitting values of the Y's.

We may now define the following *row* vectors:

m with k elements $m_{bg} - m_{0g}$ designates the correction needed to change the first (or zero-th) guess into the best b guess,

c with n elements $Y_{bi} - Y_{0i}$ designates the corresponding changes in the calculated Y_0's,

d with n elements $y_i - Y_{0i}$ designates the difference between the zero-th approximation and the observed values, and

e with n elements $y_i - Y_{bi}$ designates the error of fit for the best values.

The problem may now be stated as follows: solve for the vector m in terms of Y, M, and d so that ee' is a minimum.

From the definition of elements we see that

$$e = d - c; \tag{25}$$

also

$$c = mF_q \,, \tag{26}$$

where q designates a suitable point on the curve between m_{0g} and n_{bg}. Thus,

$$e = d - mF_q \,. \tag{27}$$

Selecting m so as to minimize ee' is the multiple regression problem, which is solved as indicated in (16, pp. 173–174). Following this procedure,

$$ee' = dd' - dF_q'm' - mF_q\, d' + mF_qF_q'm'. \tag{28}$$

Differentiating with respect to m and setting the result equal to zero gives

$$2mF_qF_q' - 2dF_q' = 0. \tag{29}$$

If the changes in parameter values are slight so that F_p is approximately equal to F_q, then substituting $M^{-1}Y$ for F_q and solving for m gives

$$m = dY'(YY')^{-1}M. \tag{30}$$

Thus, we have an approximation for the correction term m expressed in terms of known values of the trial parameters M, Y, and d, the observed and predicted values of the dependent variables.

From the correction term m we can obtain a new vector of trial parameters $m_{(k+1)g}$, which should give a set of predictions $Y_{(k+1)i}$ which is better than any of the predictions Y_{oi} to Y_{ki} previously obtained. The new vector $m_{(k+1)g}$ can be substituted in the matrix M for the vector m_{hg} giving the poorest fit, and the resulting set of $k + 1$ trial parameters used to obtain a second m vector by the use of (30).

This solution exhibits the critical requirements of the Method of False Position much more clearly than does the lengthier expression in terms of elementary algebra given by Pearson (**10**). Since YY' must have an inverse, the rank must be k; that is, the different trial values of the parameters must be such that the result from one trial is not a linear function of the results from other trials. Correspondingly, since $Y = MF$, the various trial values of the parameters must be independent of each other, since YY' will have a rank less than k if the rank of M is less than k.

Thus, to the requirements stated at the beginning of this section, we must now add that the k changes in trial parameters must be linearly independent of each other and must result in a set of linearly independent Y's. For example, changing only one parameter for each set would result in a diagonal matrix for M which would clearly have a rank of k.

The procedure is an iterative one. It may be necessary to apply it several times to find a minimum. It also is desirable to check in the vicinity of the minimum to be sure that the point found is approximately a minimum. Since the function being minimized is a quadratic, it will have only one minimum.

Furthermore, the parameter changes and changes in error must be small

enough so that a line, plane, or hyperplane is a good fit to the surface in the region being dealt with. Under these conditions the equation $Y = MF$ will be a good approximation to the surface. Changing only one parameter at a time, so that M is a diagonal matrix, would be a simple method of satisfying this requirement.

The Method of False Position with some of its extensions and limitations has been discussed by Willers (17). He also presents other methods for solving problems of the type considered here.

Predicting Scale Values of Composite Stimuli

The four laws (equations 19, 20, 21, and 24) were used to predict the scale values of the composite stimuli from the scale values of their component stimuli.

For the negative exponential law four sets of values were chosen for B^-, B^+, and C. These values were used with (24) to give sets of values for $\hat{s}_{ij}$. The parameter values were used to construct the matrix M. The values of $\hat{s}_{ij}$ gave matrix Y. Using s_{ij} and $\hat{s}_{ij}$ gave the vector d. Equation (30) is then used to find a correction which gives a fifth set of values of the parameters, which is better than any of the first four. This process is repeated until a minimum is found and tested. For a one-parameter system the process is similar but much simpler. To give a measure of goodness of fit we have presented the sum of the squares of the discrepancies as well as the sum of the absolute values of the discrepancies. These values are shown in Table 3.

It can be seen that the logarithmic and square-root laws in this case give the largest discrepancies between the actual values of the composites and the values as predicted from the single stimuli. Therefore, we shall not consider either the logarithmic or square-root laws in further detail.

Both the linear and negative exponential laws placed the zero point

TABLE 3

Four Value Laws Compared

Law	Asymptote		Additive Constant	$\Sigma\lvert s_{ij} - \hat{s}_{ij}\rvert$	$\Sigma(s_{ij} - \hat{s}_{ij})^2$
	Pos. B^+ or $A^+ + C$	Neg. B^- or $A^- + C$	C		
Negative Exponential	4.6	-1.8	+0.8	1.069	.205
Linear	∞	$-\infty$	+0.94	1.658	.434
Logarithmic	∞	$-\infty$	+1.10	2.177	.568
Square Root	∞	$-\infty$	+1.13	2.487	.815

between pork and lamb, where previous consideration showed it reasonably came.

Figure 1 shows the test for the linear value increase—the plot of the

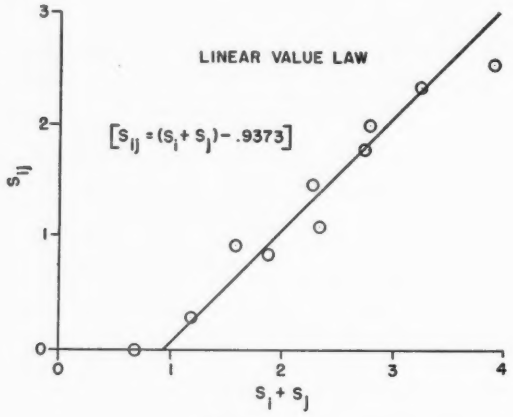


FIGURE 1

scale value of the composite against the sum of the scale values of the components. This plot gives a reasonably good fit to the line

$$\hat{s}_{ij} = (s_i + s_j) - .9373.$$

Thus, the best estimate of the zero point on the assumption of the linear law is .9373, or about .94 if we assume that $x_0 = 0$.

A more detailed analysis showing the fit for the negative exponential is shown in Figure 2. Here we have the value of tongue, .137, plotted on the abscissa, and over it the value of each of the composites with tongue—tongue and pork, tongue and lamb, tongue and beef, and tongue and steak. The same has been done for pork, lamb, beef, and steak and their composites with each of the other four stimuli.

The lines show how a negative exponential rule would fit the data, given that the zero point is at .8, that the upper asymptote is at 4.6 and the lower one at -1.8. We have a family of five lines, the upper line indicating the values v_{ij} for all composites with steak, designated S. The lower line indicates the values v_{ij} for all composites with tongue, designated T. Correspondingly, the other three lines show the values for composites with pork, lamb, and beef. These lines all converge at the two points (4.6, 4.6) and $(-1.8, -1.8)$, corresponding to the two asymptotes. It can be seen that the fit is good.

These data are not adequate to discriminate between the different

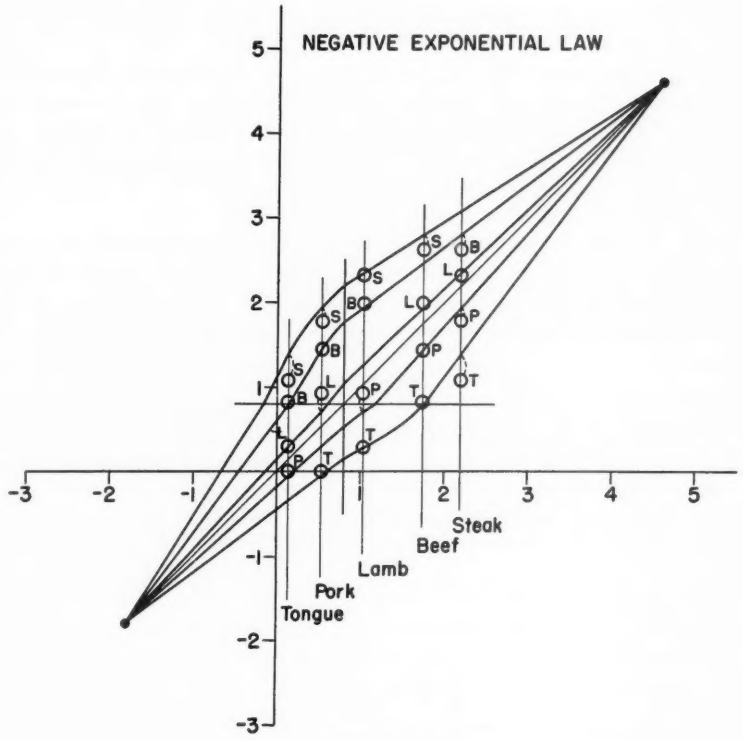


FIGURE 2

value laws even though the discrepancies are smallest for the negative exponential. More points are needed, particularly more points in the upper right quadrant, e.g., where both components are positive and hence the composite is positive. Similarly, if one has negative values, there should be more of them, so that the negative components would combine to form a number of different negative composites. It had been expected that all the values in this case would be positive and hence that five components might have been adequate.

Summary

Four different value laws have been developed: a square-root, a logarithmic, a negative exponential, and an additive law. A method has been presented for testing each law using only the scale values of components and composite stimuli determined by a psychophysical scaling method. In this case paired comparisons was used. A tentative extension of these rules has

been made to cover the case in which both negative and positive components are combined.

It has been shown that either the linear or the negative exponential law gives a good fit to limited data on food preferences. Also, it should be noted that persons do make consistent judgments about preferences for composite stimuli of the type used here, so that it is experimentally feasible to secure consistent judgments involving (i and j vs. g) or (i and j vs. g and h) in addition to the usual (i vs. j) type of choice.

Thus, we have a procedure for investigating the laws governing preferences, or value judgments, in areas where there is no readily available method for obtaining a physical measure of the amount of the commodity, and in which the usual scaling methods do not give a zero point.

REFERENCES

1. Bishop, Ruth. Points of neutrality in social attitudes of delinquents and non-delinquents. *Psychometrika*, 1940, **5**, 35-45.
2. Cauchy, Augustin Louis. Analyse algébrique, 1821. (Chapter V, Problem I.)
3. Gulliksen, H. A rational equation of the learning curve based on Thorndike's law of effect. *J. gen. Psychol.*, 1934, **11**, 395-434.
4. Gulliksen, H. Paired comparisons and the logic of measurement. *Psychol. Rev.*, 1946, **53**, 199-213.
5. Gulliksen, H. A least squares solution for paired comparisons with incomplete data. *Psychometrika*, 1956, **21**, 125-134.
6. Hamel, Georg. Eine Basis aller Zahlen und die unstetigen Lösungen der Funktionalgleichung $f(x + y) = f(x) + f(y)$. *Mathematische Annalen*, 1905, **60**, 459-462.
7. Horst, A. P. A method for determining the absolute value of a series of stimulus situations. *J. educ. Psychol.*, 1932, **23**, 418-440.
8. Mitscherlich, E. A. Das Gesetz des Minimums und das Gesetz des abnehmenden Bodenertrages. *Landwirtschaftl. Jarhb. Zsch. f. wiss. Landwirtschft*, 1909, **38**, 537-552.
9. Osgood, W. F. Advanced calculus. New York: MacMillan, 1940.
10. Pearson, Karl. On a general theory of the method of false position. *Phil. Mag.*, 1903, **4**, 658-668 (6th series).
11. Sierpiński, Waclaw. Sur l'équation fonctionelle $f(x + y) = f(x) + f(y)$. *Fundamenta Mathematicae*, 1920, **1**, 116-122.
12. Spillman, W. J. and Lang, E. The law of diminishing returns. Yonkers, N. Y.: World Book Co., 1924.
13. Thurstone, L. L. The indifference function. *J. soc. Psychol.*, 1931, **2**, 139-167.
14. Thurstone, L. L. Measurement of values. *Psychol. Rev.*, 1954, **61**, 47-58.
15. Thurstone, L. L. and Jones, L. V. The rational origin for measuring subjective values. (Unpublished manuscript.)
16. Turnbull, H. W. and Aitken, A. C. An introduction to the theory of canonical matrices. London and Glasgow: Blackie and Son, Limited, 1950.
17. Willers, F. A. (Trans. by R. T. Beyer) Practical analysis graphical and numerical methods. New York: Dover Publ. Inc., 1948.

Manuscript received 4/7/55

Revised manuscript received 10/3/55

PSYCHOMETRIKA—VOL. 21, NO. 3
SEPTEMBER, 1956

MAXIMIZING TEST BATTERY PREDICTION WHEN THE WEIGHTS ARE REQUIRED TO BE NON-NEGATIVE

JOSEPH LEV

NEW YORK STATE DEPARTMENT OF CIVIL SERVICE

A procedure is developed for computing optimum regression weights under the restriction that they be non-negative. The weights maximize, subject to the restrictions, the multiple correlation between several predictors and a criterion. A numerical example is provided.

For some purposes weights of subtests of an examination must be positive. This is true, for example, in civil service examinations when weights are announced in advance of administration of the examinations. Positive weights are needed in order to motivate candidates to perform well in the subtests. The problem considered in this paper is the computation of weights of subtests so as to maximize prediction of a criterion, with the restriction that the weights be non-negative.

The procedure to be described is similar to the method of "steepest descent" (e.g., **1**, p. 47). Actually it is more appropriate to call this a method of steepest ascent, since the weights are determined by successive addition of increments which tend to increase the multiple correlation between the predictors and the criterion.

An iterative procedure to accomplish the purpose considered in this paper was previously described in (**2**). Advantages of the present method are that the computations at each step in the iteration indicate the procedure at the next step, the termination of the procedure is clearly indicated, and the weights obtained are the best possible.

In order to write the necessary formulas, the following notation will be adopted:

X_0 = criterion
X_i = predictors ($i = 1, 2, \cdots, n$)
a_i = weights for predictors ($i = 1, 2, \cdots, n$)
$T = a_1X_1 + a_2X_2 + \cdots + a_nX_n$ = total weighted predictor score
$C_{ii} = \sum (X_i - \bar{X}_i)^2$ = sum of squares for X_i
$C_{ij} = \sum (X_i - \bar{X}_i)(X_j - \bar{X}_j)$ = sum of products for X_i and X_j
R = correlation between X_0 and T

The L_{ij} used by Wherry and Gaylord may be used here in place of the C_{ij} . The C_{ij} are smaller numbers than the L_{ij} and are sufficient for the

accuracy required. In this notation

$$R = \frac{\sum_1^n a_i C_{0i}}{\sqrt{C_{00}} \sqrt{\sum_{i,j=1}^n a_i a_j C_{ij}}}. \tag{1}$$

In (1) R may be viewed as a function of the variables a_i ; the C_{ij} are constants. It is convenient to introduce the total differential

$$dR = \frac{\partial R}{\partial a_1} da_1 + \frac{\partial R}{\partial a_2} da_2 + \cdots + \frac{\partial R}{\partial a_n} da_n . \tag{2}$$

Here dR is an increment in R, da_i are increments in the weights, and $\partial R/\partial a_i$ are partial derivatives. (2) shows how changes in weights influence changes in R. If partial derivatives have been computed for fixed values of the a_i , values of the da_i can be selected so as to make dR positive; that is, increments can be chosen for the weights so as to cause an increase in R. If increments are chosen corresponding to large values of the partial derivatives, the increase in R to the maximum can be made with few iterations. For this reason the present procedure is a method of "steepest ascent."

For computation of the partial derivatives we note that

$$\frac{\partial R}{\partial a_i} = (C_{00} \sum_{i,j} a_i a_j C_{ij})^{-1/2} [C_{0i} - k \sum a_j C_{0j}], \tag{3}$$

where $k = \sum a_j C_{0j} / \sum_{i,j} a_i a_j C_{ij}$.

In the computations to be presented, only relative values of the partial derivatives are required. Consequently, only the quantities within the square brackets are computed.

In essence the computation involves repeated trials of increments in the weights, da_i , so that each trial produces a positive increment, dR, in R with the restriction that the a_i be positive. Values of the da_i selected at any trial are determined by the values of the partial derivatives computed at the previous trial. The trials are terminated when all of the partial derivatives are either zero (or nearly zero) or definitely negative. At this stage the terminal weights a_{ti} are computed as the sums of the increments tried at the various iterations. For those variables which have zero partial derivatives the terminal weights should be positive or zero. For the variables which have negative partial derivatives, the terminal weights should be zero.

It is evident that a maximum in R has been reached at the termination point described above. For the variables having zero partial derivatives, the terminal values of the a_i provide the usual maximum in R. For the remaining variables an increase in R can be obtained only by use of negative weights.

A justification for the procedure will be discussed more fully after the computations have been described.

Computational Procedures

The computing procedures will be described first symbolically, and then in relation to a numerical application. The symbolic development which follows is outlined in Table 1.

1. Record the sums of squares and cross products of the predictor and criterion variables as a square matrix.

2. It is helpful, though not essential, to record below this matrix the correlations between the predictors and the criterion.

3. Below the matrix record the increments corresponding to the iterations in the appropriate columns. All the increments corresponding to one iteration are placed in a single row. The symbol da_{ij} will be used to denote the increment for the variable X_j in the ith iteration.

4. After m iterations the weight a_{mj} for variable X_j is the sum of the increments used to this point,

$$a_{mj} = \sum_{i=1}^{m} da_{ij} .$$

The terminal weight for X_j is the a_{tj} after all iterations have been completed.

5. The first increment is obtained for the variable, X_j , which has highest correlation with the criterion. It is an approximation of one or two digits to the ratio C_{0j}/C_{jj} . Increments for the remaining variables are ordinarily, though not necessarily, taken to be zero in the first iteration. These increments are also the weights after the first iteration. That is, $a_{1j} = da_{1j}$ for each predictor X_j .

6. The computations which lead to the next iteration are carried out in columns set up at the right of the matrix of sums of squares and cross products.

7. In the first column of this set-up the following computations are made successively:

$$\sum a_{1j}C_{1j} , \sum a_{1j}C_{2j} , \cdots , \sum a_{1j}C_{nj} , \sum a_{1j}C_{0j} ,$$

$$\sum_i \sum_j a_{1i}a_{1j}C_{ij} = \sum_i a_{1i}\left(\sum_j a_{1j}C_{ij}\right) \qquad (i, j = 1, 2, \cdots , n),$$

$$k_1 = \left(\sum a_{1j}C_{0j}\right)/\left(\sum \sum a_{1i}a_{1j}C_{ij}\right),$$

$$C_{00}R^2 = k_1 \sum a_{1j}C_{0j} . \tag{4}$$

Formula (4) provides a measure of the correlation between the weighted sum and the criterion, attained as a result of the first iteration. A similar quantity is computed at each iteration.

TABLE 1
SYMBOLIC LAY-OUT FOR COMPUTATION OF WEIGHTS, SHOWING TWO ITERATIONS

	X_1 . . . . X_n	X_0	I	II
X_1	C_{11} . . . . C_{1n}	C_{01}	$\Sigma\, a_{1j}C_{1j}$	$\Sigma\, a_{2j}C_{1j}$
.				
.				
.				
X_n	C_{n1} C_{nn}	C_{0n}	$\Sigma\, a_{1j}C_{nj}$	$\Sigma\, a_{2j}C_{nj}$
X_0	C_{01} C_{0n}	C_{00}	$\Sigma\, a_{1j}C_{0j}$	$\Sigma\, a_{2j}C_{0j}$
	r_{01} r_{0n}		$\Sigma\Sigma\, a_{1i}a_{1j}C_{ij}$	$\Sigma\Sigma\, a_{2i}a_{2j}C_{ij}$
I	da_{11} da_{1n}		$k_1 = \dfrac{\Sigma\, a_{1j}C_{0j}}{\Sigma\Sigma\, a_{1i}a_{1j}C_{ij}}$	$k_2 = \dfrac{\Sigma\, a_{2j}C_{0j}}{\Sigma\Sigma\, a_{2i}a_{2j}C_{ij}}$
II	da_{21} da_{2n}			
.			$k_1 \Sigma\, a_{1j}C_{0j}$	$k_2 \Sigma\, a_{2j}C_{0j}$
.				
.				
t	da_{t1} da_{tn}		$C_{01} - k_1 \Sigma\, a_{1j}C_{1j}$	$C_{01} - k_2 \Sigma\, a_{2j}C_{1j}$
	.		.	
			.	
			.	
Total	a_{t1} a_{tn}			
Regression weight	$k_t a_{t1}$ $k_t a_{tn}$		$C_{0n} - k_1 \Sigma\, a_{1j}C_{nj}$	$C_{0n} - k_2 \Sigma\, a_{2j}C_{nj}$

The succeeding computations have as their purpose an evaluation of the partial derivatives exhibited in (2). We compute

$$C_{01} - k_1 \sum a_{1i}C_{1i} , \cdots , C_{0n} - k_1 \sum a_{1i}C_{ni} .$$

For the first column some simplification of the above computations is possible. However, the saving is slight and the formulas are stated in full generality as they apply to all columns. Note that the partial derivative which corresponds to the variable having positive weight becomes zero, or nearly so.

8. The increment for the second iteration is based on the values of the partial derivatives obtained from the first iteration. Ordinarily we choose the partial derivative which has the highest positive value and select an increment for the corresponding variable. If this variable is X_h , then the increment da_{2h} is chosen so that

$$C_{0h} - \sum_i a_{2i}C_{hi} = 0$$

approximately. Here the a_{2i} are the weights after the second iteration. *Increments for more than one predictor may be chosen at any iteration if this seems appropriate.*

9. To perform the computations for the second iteration we take advantage of the relationship

$$\sum_i a_{2i}C_{ij} = \sum_i a_{1i}C_{ij} + \sum_i (da_{2i})C_{ij} .$$

However, this simplification is not appropriate for the computation of $\sum \sum a_{2i}a_{2j} C_{ij}$. The remaining computations are carried out as described for the first iteration.

10. The iterations are continued until the partial derivatives are decidedly negative, or approximately zero, and the quantity $k_m \sum a_{mi}C_{0i}$ seems to have attained a maximum.

Details of a numerical example are shown in Table 2. The first positive increment, $da_{12} = 0.6$, is an approximation to the quotient 1790/3246. This increment is also a_{12} . The computations based on this increment are shown in the column headed I. The first seven entries in this column are products of the increment .6 and the cross products under X_2 as .6 (2802) = 1681, $\cdots$, .6 (1790) = 1074. The eighth entry is .6 (1948) = 1169; k_1 is 1074/1169 = .9187; $k_1 \sum a_{1i}C_{0i} = C_{00}R^2$ is given as .9187 (1074) = 986.7. The last six entries in this column indicate the relative values of the partial derivatives. The greatest of these is the one related to X_6 and has the value 2520 − (.9187) (1654) = 1000. This value of the partial derivative suggests the positive increment $da_{26} = 0.1$ for the second iteration. Computations in the

TABLE 2

ITERATIVE COMPUTATION OF WEIGHTS

	X_1	X_2	X_3	X_4	X_5	X_6	X_0	Formula		I	II	III	IV	V
X_1	7124	2802	1046	1083	320	1794	1254	$\Sigma a_j C_{1j}$		1681	1860	1878	1887	1892
X_2	2802	3246	1281	1617	570	2757	1790	$\Sigma a_j C_{2j}$		1948	2224	2252	2266	2274
X_3	1046	1281	2025	1347	1713	1581	534	$\Sigma a_j C_{3j}$		769	927	943	951	958
X_4	1083	1617	1347	3752	2499	2206	979	$\Sigma a_j C_{4j}$		970	1191	1213	1224	1243
X_5	320	570	1713	2499	9518	-1263	-65	$\Sigma a_j C_{5j}$		342	216	203	197	209
X_5	1794	2757	1581	2206	-1263	13380	2520	$\Sigma a_j C_{6j}$		1654	2992	3126	3193	3204
X_0	1254	1790	534	979	-65	2520	12438	$\Sigma a_j C_{0j}$		1074	1326	1351	1364	1369
r	.1332	.2817	.1063	.1434	-.0060	.1954		$\Sigma\Sigma a_i a_j C_{ij}$		1169	1634	1695	1727	1739
								k		.9187	.8115	.7971	.7898	.7872
I	0	.6	0	0	0	0		$k \Sigma a_j C_{0j}$		986.7	1076.0	1076.9	1077.3	1077.7
II	0	0	0	0	0	.1		$C_{01} - k \Sigma a_j C_{1j}$		-291	-256	-243	-236	-235
III	0	0	0	0	0	.01		$C_{02} - k \Sigma a_j C_{2j}$		0	-16	-5	0	0
IV	0	0	0	0	0	.005		$C_{03} - k \Sigma a_j C_{3j}$		-173	-219	-218	-217	-220
V	0	0	0	.005	0	0		$C_{04} - k \Sigma a_j C_{4j}$		-88	12	12	12	1
Total	0	.6	0	.005	0	.115		$C_{05} - k \Sigma a_j C_{5j}$		-379	-240	-227	-221	-230
Regression weight	0	.4723	0	.0039	0	.0905		$C_{06} - k \Sigma a_j C_{6j}$		1000	90	28	-2	-2

column headed II will be given in full. Notice that

$$
\begin{aligned}
1681 + .1(1794) &= 1860 \\
1948 + .1(2757) &= 2224 \\
769 + .1(1581) &= 927 \\
970 + .1(2206) &= 1191 \\
342 + .1(-1263) &= 216 \\
1654 + .1(13380) &= 2992 \\
1074 + .1(2520) &= 1326 \\
.6(2224) + .1(2992) &= 1634 \\
1326/1634 &= .8115 \\
.8115\ (1326) &= 1076.0 \\
1254 - (.8115)\ (1860) &= -256 \\
1790 - (.8115)\ (2224) &= -16 \\
534 - (.8115)\ (927) &= -219 \\
979 - (.8115)\ (1191) &= 12 \\
-65 - (.8115)\ (216) &= -240 \\
2520 - (.8115)\ (2992) &= 90
\end{aligned}
$$

Iterations and computations continue as shown in Table 2. At iteration 5 the process stops because the partial derivatives for variables X_2 , X_4 , X_6 are nearly zero and the remaining partial derivatives are negative. Note also the stability of $C_{00}R^2$ over the last three iterations. The result is a multiple R of .2994, with regression weights 0, .4723, 0, .0039, 0, .0905.

All the increments tried in this numerical example have been positive. In some circumstances negative increments may be tried. This is true when a positive weight has been too large as shown by a negative partial derivative for a variable having positive weight.

An evaluation of the method is called for at this point. One may ask whether the weights obtained by this method actually provide the highest possible R under the restriction of non-negative weights. Before dealing with this question it is necessary to consider whether there exists a set of positive values a_i which yield a maximum for R.

One way of demonstrating that a maximum actually exists is to note first that under ordinary conditions a set of weights maximizing R can be found for any selection of variables out of the n given variables. Some of the selections of variables will give a maximum R with non-negative weights. There are a finite number of these selections. Consequently, we may choose that one set of variables which provides the largest maximum R under the required condition. Weights obtained in this way are the desired weights. Thus, the existence of a maximum is demonstrated. In exceptional circumstances it may happen that all weights are negative. This will be shown by the computation.

The existence of a maximum R under the condition of non-negative

weights having been established, it is necessary to demonstrate that the procedure described in this paper actually leads to this maximum. This will be demonstrated with the aid of a theorem.

The necessary and sufficient condition for the set of weights $a_1, a_2, \cdots, a_n$ *to maximize* R*, with the provision that none of the weights be negative, is that the partial derivatives* $\partial R/\partial a_i$ *be zero for variables* X_i *with positive weights, and negative or zero for variables* X_i *with zero weights.*

Necessity: This has already been discussed. If a set of non-negative values of the a_i maximizing R has been attained then the condition implies that R can be further increased only by changing from zero to negative some of the a_i, for variables having negative partial derivatives. Any change in values of the a_i for variables which have zero partial derivatives will decrease R. Thus necessity is demonstrated.

Sufficiency: To demonstrate sufficiency, suppose that the procedure of this paper has provided a set of a_i satisfying the conditions of the theorem with corresponding R. Suppose now that there is another set of values of the a_i also satisfying the conditions of the theorem with a value of $R' > R$. But this is impossible, for some of the partial derivatives for the first set of a_i should then be positive. Thus sufficiently is also demonstrated.

Since the procedure described in this paper leads to zero or negative partial derivatives, the procedure provides a unique maximum under the restriction to non-negative weights.

REFERENCES

1. Householder, A. S., Principles of numerical analysis. New York: McGraw-Hill, 1953.
2. Wherry, R. J. and Gaylord, R. H. Test selection with integral gross score weights. *Psychometrika*, 1946, 11, 173-183.

Manuscript received 5/27/55

Revised manuscript received 10/17/55

PSYCHOMETRIKA—VOL. 21, NO. 3
SEPTEMBER, 1956

THE MATCHING PROBLEM*

EDGAR J. GILBERT

HARVARD UNIVERSITY†

Tables of the exact distributions of number of matches are given for small decks having the same number of cards in each suit. Several approximate distributions are considered for use with larger decks, and some indication of the goodness of the approximations is given.

Many psychological experiments involve testing the ability of a person (or a method) to classify certain objects into categories. The matching-problem technique to be described in this paper provides a mathematical model which may be used to calculate significance levels for such experiments.

The example of the matching problem which is most often cited is the work done in certain of the early ESP (extra-sensory perception) experiments. A deck of 25 cards containing 5 each of 5 different figures (circle, cross, wave, square, and star) was shuffled into a random order by the experimenter. The subject was given a second deck of the same composition and asked to arrange it in the same order as the hidden deck of the experimenter. Then the two decks were compared. If the first card of the subject's deck was of the same kind (circle, cross, etc.) as the first card of the experimenter's deck, the subject scored a "match." Then the second card in each deck was examined, and so on through the two decks. The subject's ability was scored according to the total number of correct matches, as compared to the number to be expected by chance alone.

For another example, suppose a handwriting expert claims he can tell a person's profession by examining a sample of his handwriting. To test his ability he is given 10 samples, 2 written by doctors, 4 by lawyers, and 4 by teachers. He is told which professions are represented but not how many samples are from each profession. If we number the samples from 1 to 10, placing under each number the true profession and then the expert's guess, we might get the following result:

Sample Number:	1	2	3	4	5	6	7	8	9	10
True Profession:	L	D	L	T	D	T	L	T	T	L
Expert's Guess:	L	D	T	T	D	T	D	L	D	D

*The calculations for this paper were done while the writer was working on a project sponsored by funds from the Ford Foundation. The writer wishes to express his gratitude to Dr. Frederick Mosteller for his constant help and encouragement during the writing of this paper.

†Now at University of California, Berkeley.

In order to evaluate the expert's ability on the basis of this test, we might think of the two rows "True Profession" and "Expert's Guess" as being the arrangements of a target deck and a call deck, respectively, and imagine that the expert was attempting a kind of matching problem similar to that of the ESP experiment described above. We notice that he called correctly cards numbered 1, 2, 4, 5, and 6, for a total of 5 correct matches. However, we see that this case is different from the ESP case in that the deck used by the expert does not have the same composition as that used by the experimenter; so we must allow for this in calculating the probability of his getting the 5 matches by chance. The probabilities for this particular example will be worked out in the last section of the paper, but first the more easily calculated cases where both decks are identical will be considered.

In general, the model will consist of two decks of cards: a target deck, the suits of which correspond to the kinds of objects being classified, and a call deck whose suits correspond to the categories used by the subject in his classification. We imagine that the cards in the target deck have some arbitrary arrangement and calculate the probability distribution of the number of correct matches between cards in the two decks on the assumption that the call deck was arranged by random shuffling. If the subject has any ability, scores as high as or higher than his will have low probability.

Several variations of the problem arise, depending on the composition of the call deck. If the person to be tested is informed of the categories to be used and the number of objects in each category, and is forced to use this information, then the call deck is identical with the target deck in composition. If he is not so informed or not forced to use the information, he may classify the objects in such a way that his call deck is different from the target deck. In this case, his ability to choose categories is not perfect. In order to use the matching problem technique, the calculation of the probability of various numbers of matches must be based on the composition of the call deck actually used. It is always possible to assume that both decks have the same suits, however, by using the trick of saying a suit which is not represented actually is present, but with zero cards in it.

In this paper, tables of the exact distributions of numbers of matches for small decks having the same number of cards in each suit will be given. Several methods of obtaining approximate distributions for larger decks will be indicated. The problems of decks with different numbers of cards in different suits, and of non-identical decks, will be discussed.

Exact Distributions

The simplest form of the matching problem is the case where both the target deck and the call deck have n distinct cards, i.e., where any given card of the call deck can match one and only one card in the target deck. This is a special form of a problem which has a long history in the mathe-

matical literature. For example, Feller (**3**) gives a formula for calculating the probability distribution of the number of matches in this case and gives tables of the calculated values for decks of size 3, 4, 5, 6, and 10. He shows that for deck sizes larger than 10, the values are, to five decimal places, the same as for the Poisson distribution with unit mean.

A special case of the matching problem arises when there are only two suits in each deck and each suit has c cards in it, making a total of $2c$ cards in the deck. There is no possible arrangement of the call deck which will give an odd number of matches, while the number of arrangements giving h matches, when h is an even integer, is $\binom{c}{h/2}^2$, the square of the number of combinations of c things taken $h/2$ at a time. Since this is the square of a binomial coefficient, it is a simple matter to look up the binomial coefficients corresponding to $h = 0, 2, 4, \cdots 2c$, square them, and add the results to get the total number of ways of arranging the deck. Then divide the number of arrangements which give h matches by the total number of arrangements to get the probability of exactly h matches. For example, if each suit has 2 cards in it, then there are $\binom{2}{0}^2 = 1$ way of getting 0 matches, $\binom{2}{1}^2 = 4$ ways of getting 2 matches, and $\binom{2}{2}^2 = 1$ way of getting 4 matches. So the probability of exactly h matches when $h = 0, 1, 2, 3, 4$ is $1/6, 0, 4/6, 0, 1/6$.

The next easiest case to calculate is that of two identical decks, each having s suits of c cards per suit. It is this situation, where the probability of exactly h matches is a function of the three numbers: s, the number of suits; c, the number of cards in each suit; and h, the number of matches, which will be the primary concern in this paper. Denote this probability by $m(s, c; h)$ and write the probability of h or more matches (which is the sum of the probabilities for h, $h + 1$, $h + 2$, etc.) as $M(s, c; h)$. It is this latter probability which is usually wanted in finding the significance of experimental results. We usually want to know the probability of a subject's doing as well as or better than his result by chance alone.

Even in this case, few tables have been published giving the exact probabilities. Huntington (**8**) gives tables of $m(3, 3; h)$ and $m(4, 4; h)$. Greville (**5**) gives a table of $m(5, 5; h)$, the probabilities for the ESP experiment referred to earlier. Greenwood (**4**) gives a table of estimated values of $m(4, 13; h)$ for values of h from 0 to 7, where both decks are the ordinary 52-card bridge decks. There may be other tables available in the literature; but it was thought that a small collection of tables in one place might be useful.

Greville (**6**) derived a formula for the matching problem distribution which seemed to the writer to be more adapted to calculation of exact probabilities than some others in the literature. His formula was for a more general case than that we are now considering and will be used in its general form in

TABLE 1

Exact Distributions. Probability of h or More Matches

h	M(2, 1;h)	M(3, 1;h)	M(4, 1;h)	M(5, 1;h)	M(6, 1;h)	M(7, 1;h)	M(8, 1;h)	M(9, 1;h)	M(10, 1;h)	M(11, 1;h)
0	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
1	.50000	.66667	.62500	.63333	.63194	.63214	.63212	.63212	.63212	.63212
2	.50000	.16667	.29167	.25833	.26528	.26409	.26426	.26424	.26424	.26424
3		.16667	.04167	.09167	.07778	.08075	.08023	.08031	.08030	.08030
4			.04167	.00833	.02222	.01825	.01912	.01897	.01899	.01899
5				.00833	.00139	.00436	.00350	.00369	.00366	.00365
6					.00139	.00020	.00072	.00056	.00060	.00059
7						.00020	.00002	.00010	.00008	.00008
8							.00002	.00000	.00001	.00001

h	M(2, 2;h)	M(3, 2;h)	M(4, 2;h)	M(5, 2;h)	M(6, 2;h)	M(7, 2;h)	M(8, 2;h)	M(9, 2;h)	M(10, 2;h)	M(11, 2;h)
0	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
1	.83333	.88889	.88214	.87870	.87632	.87462	.87335	.87237	.87158	.87094
2	.83333	.62222	.61548	.60991	.60689	.60483	.60334	.60220	.60131	.60060
3	.16667	.32222	.32341	.32451	.32406	.32383	.32369	.32360	.32354	.32350
4	.16667	.14444	.13294	.13262	.13450	.13580	.13676	.13749	.13806	.13853
5		.01111	.03532	.04179	.04396	.04538	.04640	.04716	.04776	.04824
6		.01111	.00992	.01033	.01144	.01226	.01285	.01330	.01365	.01394
7			.00040	.00177	.00237	.00270	.00295	.00313	.00328	.00340
8			.00040	.00036	.00041	.00049	.00056	.00062	.00067	.00071
9				.00001	.00005	.00007	.00009	.00010	.00012	.00013
10				.00001	.00001	.00001	.00001	.00002	.00002	.00002

h	M(2, 3;h)	M(3, 3;h)	M(4, 3;h)	M(5, 3;h)	M(6, 3;h)	M(7, 3;h)	M(8, 3;h)	M(2, 4;h)	M(3, 4;h)	M(4, 4;h)
0	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
1	.95000	.96667	.96258	.96016	.95852	.95734	.95645	.98571	.99001	.98813
2	.95000	.83810	.82875	.82266	.81875	.81603	.81402	.98571	.93737	.93005
3	.50000	.61310	.60010	.59456	.59113	.58882	.58715	.75714	.80646	.79355
4	.50000	.35417	.35404	.35335	.35313	.35301	.35295	.75714	.60052	.58930
5	.05000	.16131	.16724	.17154	.17407	.17578	.17702	.24286	.37117	.37156
6	.05000	.04881	.06322	.06808	.07113	.07319	.07467	.24286	.18831	.19645
7		.01667	.01864	.02215	.02425	.02569	.02674	.01428	.07472	.08658
8		.00060	.00461	.00592	.00694	.00766	.00820	.01428	.02485	.03174
9		.00060	.00073	.00130	.00167	.00195	.00217		.00511	.00962
10			.00015	.00023	.00034	.00043	.00050		.00141	.00242
11			.00000	.00004	.00006	.00008	.00010		.00003	.00049
12			.00000	.00000	.00001	.00001	.00002		.00003	.00009
13										.00001

13			.00000	.00000	.00001	.00001	.00002		.00003	.00009
										.00001

TABLE 1 (Continued)

h	M(5,4;h)	M(2,5;h)	M(3,5;h)	M(4,5;h)	M(5,5;h)	M(2,6;h)	M(3,6;h)	M(2,7;h)	M(3,7;h)	M(2,8;h)
0	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
1	.98693	.99603	.99702	.99624	.99571	.99892	.99911	.99971	.99974	.99992
2	.92558	.99603	.97700	.97286	.97026	.99892	.99186	.99971	.99720	.99992
3	.78646	.89682	.91278	.90271	.89690	.95996	.96331	.98543	.98532	.99495
4	.58385	.89682	.78150	.76776	.76033	.95996	.89136	.98543	.94969	.99495
5	.37138	.50000	.59102	.58113	.57600	.71645	.76155	.85693	.87266	.93403
6	.20090	.50000	.38388	.38408	.38406	.71645	.58399	.85693	.74532	.93403
7	.09232	.10397	.21077	.21933	.22380	.28355	.39331	.50000	.57839	.69036
8	.03609	.10397	.09571	.10764	.11374	.28355	.22923	.50000	.40070	.69036
9	.01202	.00397	.03624	.04528	.05043	.04004	.11416	.14306	.24472	.30964
10	.00342	.00397	.01074	.01630	.01953	.04004	.04838	.14306	.13051	.30964
11	.00083		.00281	.00500	.00662	.00108	.01709	.01457	.06040	.06597
12	.00017		.00043	.00131	.00196	.00108	.00513	.01457	.02404	.06597
13	.00003		.00010	.00029	.00051		.00120	.00029	.00823	.00505
14			.00000	.00005	.00012		.00026	.00029	.00237	.00505
15				.00001	.00002		.00003		.00059	.00008
16							.00001		.00011	.00008
17									.00002	

h	M(2,9;h)	M(2,10;h)	M(2,11;h)	M(2,12;h)
0	1.00000	1.00000	1.00000	1.00000
1	.99998	.99999	1.00000	1.00000
2	.99998	.99999	1.00000	1.00000
3	.99831	.99945	.99983	.99995
4	.99831	.99945	.99983	.99995
5	.97166	.98849	.99554	.99834
6	.97166	.98849	.99554	.99834
7	.82653	.91055	.95694	.98044
8	.82653	.91055	.95694	.98044
9	.50000	.67186	.80257	.88983
10	.50000	.67186	.80257	.88983
11	.17347	.32814	.50000	.65786
12	.17347	.32814	.50000	.65786
13	.02834	.08945	.19743	.34214
14	.02834	.08945	.19743	.34214
15	.00169	.01151	.04305	.11017
16	.00169	.01151	.04305	.11017
17	.00002	.00055	.00446	.01956
18	.00002	.00055	.00446	.01956

the last section of the paper. The following special case of Greville's formula was used to calculate the values given in Table 1: ($n = sc$ = size of deck)

$$m(s, c; h) = \frac{1}{n!} \sum_{i=h}^{n} (-1)^{i-h} \binom{i}{h} (n - i)! H_i , \tag{1}$$

where H_i is the coefficient of x^i in the expansion of

$$\left(\sum_{j=0}^{c} \frac{c!c!x^j}{j!(c - j)!(c - j)!} \right)^s . \tag{2}$$

Since it is usually the probability of h or more matches which is wanted in checking experimental data, Table 1 lists $M(s, c; h) = \sum_{i=h}^{n} m(s, c; i)$. The calculations were carried out to exact integers before dividing by $n!$, the divisions made to 6 decimals, and the values of $M(s, c; h)$ rounded to five decimals after summing. The table is arranged first by groups in order of increasing c, then within groups in order of increasing s. Within each group it will be noticed that the distributions seem to approach a limiting distribution—the Poisson—but that the approach is slower for groups with larger c. It will also be noticed that the first distribution in each group [that for $M(2, c; h)$] has the entry for odd h equal to the entry for the following even h. This is due to the fact mentioned earlier, that the probability in this case of exactly h matches, when h is odd, is zero.

The values $M(5, 5; h)$ given in Table 1 were obtained by summing to five decimals the values given to 7 decimals by Greville (**5**). All other values in Table 1 were calculated independently by the writer, using the tables of Feller and Huntington as a check where there were duplications.

Approximate Distributions

Since the calculation of exact probabilities gets very tedious for large decks, it would be desirable to be able to use distributions which either are already available, or are easier to calculate, to approximate significance levels for the number of matches. As has already been mentioned, Feller's tables (**3**) show that a Poisson distribution is satisfactory for the approximation of $m(s, 1; h)$, when s is only moderately large. A study of the distributions within each group with the same value of c (Table 1) leads one to suspect that this is true also of groups other than $c = 1$; but as c gets large, the rate of approach to Poisson with increasing s is much slower.

In order to find more accurate approximations, we need to know more about the moments of the matching problem distribution. We shall continue throughout this section to work with the relatively simple case of identical decks with the same number of cards in each suit. For this case, the first four moments were derived by means of a rather ingenious use of determinants by Olds (**9**). If ν is the mean and μ_i the ith moment about the mean,

then in terms of c, the number of cards per suit, and $n = sc$, the size of the deck, we have

$$\begin{aligned}
\nu &= c \\
\mu_2 &= c(n - c)/(n - 1) \\
\mu_3 &= c(n - c)(n - 2c)/(n - 1)(n - 2) \\
\mu_4 &= \frac{c(n - c)}{(n - 1)(n - 2)(n - 3)} [(n - 2c)(n - 3c)(3c + 1) \\
&\qquad + (c - 1)(12nc - n - 18c^2 - 6c)].
\end{aligned} \tag{3}$$

Anderson (**1**) shows that as the deck size is increased while keeping the proportion of cards in each suit fixed, the number of matches is asymptotically normally distributed. So if n is large enough, the mean and variance of the matching problem distribution can be calculated from formulas (3), and the desired values found in a table of the normal distribution. Unfortunately, it is difficult to say precisely when n is large enough. In many cases, the matching problem distribution is asymmetrical; to the extent that this is so, the normal will give a bad fit. Although the normal is not as close an approximation as some of the distributions considered later in this section, it can be seen from the examples in Table 2 that it is sufficiently close for some purposes; it has the advantage of being available without too much calculation. An example of the calculations involved in fitting a normal to a matching problem is worked out in the last section of this paper.

Hamilton (**7**) noted that the matching problem distribution was somewhat like a binomial distribution with $p = 1/s$ and $n = sc$, and suggested that writing the mean of the matching problem as np and variance as $npq[n/(n - 1)]$ made the similarity apparent. It can be seen that the binomial with $p = 1/s$ has the same mean as the matching problem, and that the ratio of the two variances approaches 1 as n gets large. This is also the case with the third moment about the mean. It is $npq(2q - 1)$ for the binomial, and we can rearrange the formula for μ_3 in (3) above—substituting np for c where it appears and setting $q = 1 - p$—to make the third moment of the matching problem look like $npq(2q - 1)n^2/(n - 1)\ (n - 2)$.

The binomial is suggested as a good approximating distribution for another reason—extensive tables of the binomial are available (**11**). Since the binomial is a two-parameter distribution, we can in many cases find a tabled distribution which has both mean and variance equal to that of the matching problem distribution. In this case, we shall not look up the same values of n and p used in the last paragraph; we shall treat n' and p' simply as two parameters, using the prime to show that these are different numbers. If we replace n in (3) by sc and equate the means and variances of the match-

TABLE 2

Approximate Distributions

G-C is Gram-Charlier approximation; G-CB is the binomial modification of Gram-Charlier. Values of M(4, 13;h) and its 2d order G-C are taken from Greenwood (4), summed to four decimal places. The binomial approximation to M(4, 13;h) was chosen by method of text to have same mean and variance as exact distribution.

h	Exact M(8, 2;h)	Poisson $\nu = 2$	Binomial n=25;p=.08	4th Order G-C
0	1.00000	1.00000	1.00000	1.00000
1	.87335	.86466	.87564	.87330
2	.60334	.59399	.60528	.60340
3	.32369	.32332	.32317	.32361
4	.13676	.14288	.13509	.13671
5	.04640	.05265	.04514	.04643
6	.01285	.01656	.01229	.01287
7	.00295	.00435	.00277	.00295
8	.00056	.00110	.00052	.00055
9	.00009	.00024	.00008	.00009
10	.00001	.00005	.00001	.00001

h	Exact M(8, 3;h)	Poisson $\nu = 3$	Binomial n=30;p=.10	4th Order G-CB
0	1.00000	1.00000	1.00000	1.00000
1	.95645	.95021	.95761	.95646
2	.81402	.80085	.81630	.81400
3	.58715	.57681	.58865	.58712
4	.35295	.35277	.35256	.35295
5	.17702	.18474	.17549	.17704
6	.07467	.08392	.07319	.07468
7	.02674	.03351	.02583	.02674
8	.00820	.01191	.00778	.00820
9	.00217	.00380	.00202	.00217
10	.00050	.00110	.00045	.00050
11	.00010	.00029	.00009	.00010
12	.00002	.00007	.00002	.00002

h	Exact M(4, 13;h)	2nd Order G-C	Binomial n=68;p=1/17	Normal
0	1.00000	1.00000	1.00000	.990
1	.9838	.9839	.9839	.964
2	.9149	.9149	.9149	.902
3	.7707	.7705	.7708	.780
4	.5725	.5723	.5726	.602
5	.3712	.3712	.3713	.398
6	.2099	.2103	.2102	.210
7	.1047	.1045	.1045	.098

ing problem distribution and the binomial,

$$\nu = c = n'p'$$

$$\mu_2 = c^2(s - 1)/(sc - 1) = n'p'(1 - p'),$$

a pair of simultaneous equations which we can solve for n' and p' in terms of

s and c. For solutions

$$n' = c(sc - 1)/(c - 1)$$
$$p' = (c - 1)/(sc - 1). \tag{4}$$

Note that n' is not the size of the deck but is just a parameter for which we have solved. It may or may not be an integer. If it is, and p' is a value which can be found in a table, then the binomial distribution with parameters n' and p' will have the same mean and same variance as the matching problem distribution with s suits of c cards each. If n' is not an integer, the integer nearest to n' can be used, call it n'', and using $p'' = c/n''$ gives a binomial with the correct mean and very nearly the correct variance. The writer would guess, based on trying a few cases, that for decks larger than 25, one should be able to fit a binomial by this method to about three decimal places. Table 2 shows some examples of the use of the binomial as an approximation.

This approach suggests looking for known, or easily calculated, distributions which have the first two, three, or more moments in common with the matching problem distribution. Greenwood (**4**), following a suggestion by Mantel, used the Gram-Charlier series type B as an approximation. The Gram-Charlier uses the Poisson as a first approximation to the desired distribution; suitable multiples of the first, second, and successive differences of the Poisson are added to correct for the mean, second moment, etc. If $p(h)$ is the Poisson probability of h, then the first difference, $\Delta p(h)$, is defined to be $p(h) - p(h - 1)$. The second difference, $\Delta^2 p(h)$, is the first difference of the first difference, i.e., $\Delta p(h) - \Delta p(h - 1)$, and so on for higher differences. The reader will note that $\Delta p(0)$ is not well defined yet, since we must know what value to use for $p(-1)$—this is not ordinarily found in a table of the Poisson distribution. For purposes of the Gram-Charlier series, $p(h)$ is defined and equal to zero for all negative values of h. We are now ready to define the Gram-Charlier series. Where $m(h)$ is the distribution being approximated,

$$m(h) = p(h) + a_1\Delta p(h) + a_2\Delta^2 p(h) + a_3\Delta^3 p(h) + \cdots . \tag{5}$$

In principle, this is an infinite series; in practice, only the first few terms are used. The constants a_i appearing in (5) are determined by the moments of the matching problem and of the particular Poisson distribution used. The Poisson is usually chosen so that it has the same mean as the distribution being approximated; this makes the value of a_1 zero. If μ_i is the ith moment about the mean of the matching problem distribution and m_i the corresponding moment for the Poisson, the next three constants for the Gram-Charlier type B are given by

$$2a_2 = \mu_2 - m_2$$
$$6a_3 = 6a_2 - (\mu_3 - m_3) \tag{6}$$
$$24a_4 = 36a_3 - 2a_2(6m_2 + 7) + (\mu_4 - m_4).$$

As an example, suppose we wish to find an approximate value for $m(7, 2; 3)$, the probability of exactly three matches when both decks have seven suits of two cards per suit. (From Table 1 we see that the correct value is .32383 − .13580 = 0.18803.) As a first approximation, we look up the Poisson with mean 2, and find $p(3) = 0.18045$. This value is 0.00758 too low. For the next approximation, we need to know the variances and the value of $\Delta^2 p(3)$:

$$\mu_2 = c(n - c)/(n - 1) = 2(14 - 2)/(14 - 1) = 24/13$$

$$m_2 = 2$$

$$a_2 = \tfrac{1}{2}(24/13 - 2) = -1/13$$

$$\Delta p(3) = p(3) - p(2) = 0.18045 - 0.27067 = -0.09022$$

$$\Delta p(2) = p(2) - p(1) = 0.27067 - 0.27067 = 0.00000$$

$$\Delta^2 p(3) = \Delta p(3) - \Delta p(2) = -0.09022$$

$$a_2\Delta^2 p(3) = (-1/13)(-0.09022) = 0.00694.$$

Therefore, we have as a second approximation $p(3) + a_2\Delta^2 p(3) = 0.18739$, which is only 0.00064 too low. If we wished a still closer approximation, we could calculate the terms involving a_3 and a_4 . An example of a fourth-order approximation is given in Table 2.

If the bionomial is used as a first approximation, higher-order approximations may be obtained by a modification of the Gram-Charlier series which uses the binomial instead of the Poisson. The formulas (5) and (6) are valid for this modification, provided that we re-interpret $p(h)$ and m_i to be the probability and moments of the binomial. It has been the writer's experience that the Poisson works well as a starting distribution when c is 2 or 3, and that the binomial gives a better fitting curve when c is more than 3. For decks of 24 or more cards, when all suits have the same number of cards, truncating the Gram-Charlier series (or the binomial modification) after terms which involve a_4 gives an approximation correct to about 0.00003 or less. For a more detailed description of the Gram-Charlier series and a derivation of the constants involved see Rietz (**10**).

More Complicated Decks

If the target deck does not have the same composition as the call deck, or if the suits do not all have the same number of cards, the estimation of significance levels may be more complicated. In many cases, the first complication can be avoided by the simple device of making sure the person who is attempting the classification knows what the categories are and how many objects are in each category. When it is possible to design the experiment so that this is done, several advantages are gained. The calculations for each

subject or trial are simpler for identical decks. Also, if each subject selects his own categories, there may be a different set of calculations necessary for each subject; whereas if all subjects use the same deck, only one distribution need be calculated. Finally, while the writer knows of no investigations of the power of this type of test, it seems clear that the subject has a better chance of showing his ability, if he has any, when both decks are the same. Since, in using the matching problem technique, the calculations are based on the decks actually used, there is no way to give credit for choosing the proper categories.

The calculations are also somewhat more complicated if different categories contain different numbers of objects. However, in many cases this is not a part of the experimental design which can be changed. For small decks, the exact distribution may have to be calculated. If so, the following procedure, taken from Greville (6), may be used. (This also works for calculating non-identical decks.) Suppose the two decks have the following composition:

	Suit	1	2	3	$\cdots$	s	Total
Cards Per Suit	Call Deck	m_1	m_2	m_3		m_s	n
	Target Deck	n_1	n_2	n_3		n_s	n

(7)

Let M_i be the smaller of m_i and n_i for each $i = 1, 2, \cdots, s$. Then the probability of exactly h matches, $m(h)$, is given by

$$m(h) = \frac{1}{n!} \sum_{j=h}^{n} (-1)^{j-h} \binom{j}{h} (n - j)! H_j , \tag{8}$$

where H_j is the coefficient of x^j in the expansion of

$$\prod_{i=1}^{s} \left[\sum_{k=0}^{M_i} \frac{m_i! n_i! x^k}{k!(m_i - k)!(n_i - k)!} \right].$$

As an example, suppose we wish to calculate the distribution of matches in the case of the handwriting expert given at the beginning of the paper. The two decks have the composition:

	Suit	1	2	3	Total
Cards Per Suit	Call Deck	5	2	3	10
	Target Deck	2	4	4	10

(9)

We notice one peculiar feature of non-identical decks at this point. Although both decks have 10 cards each, only 7 matches are possible: 2 in the first

suit, 2 in the second, and 3 in the third. Thus, we will expect to get non-zero values of $m(h)$ only for $h = 0, 1, \cdots, 7$.

First expand the generating function (9) to get H_i :

$$(1 + 10x + 20x^2)(1 + 8x + 12x^2)(1 + 12x + 36x^2 + 24x^3)$$
$$= 1 + 30x + 364x^2 + 2296x^3 + 8064x^4 + 15648x^5 + 15360x^6 + 5760x^7.$$

Then calculate the factor $Q_j = (n - j)!H_j$ for each $j = 0, 1, \cdots, 7$. Next calculate $n!m(h) = \sum_{j=h}^{7} (-1)^{j-h}\binom{j}{h}Q_j$, for each $h = 0, 1, \cdots, 7$, where $\binom{j}{h}$ is a binomial coefficient, which can be obtained from tables. With a table of binomial coefficients and the values of Q_j, which we have calculated, this sum can be obtained rather quickly on a desk calculator. Finally, divide each value by $n!$ to get $m(h)$, the probability of exactly h matches. In order to get the probability of h or more matches, add $\sum_{i=h}^{7} m(i)$, getting the values listed in Table 3.

TABLE 3

Handwriting Expert Example

h	Exact	Normal	Binomial
0	1.00000	0.9922	1.00000
1	.96984	.9582	.9718
2	.84762	.8508	.8507
3	.62064	.6368	.6172
4	.35556	.3632	.3504
5	.15238	.1492	.1503
6	.04444	.0418	.0474
7	.00952	.0078	.0106

To approximate this distribution, using either the normal or the binomial, first calculate the mean and variance. For this purpose use a formula derived by Battin (2). (Battin gives a brief review of the mathematical literature on the matching problem prior to 1942 and lists an extensive bibilography.) He uses the technique of generating functions to get formulas applicable to decks of arbitrary composition. He also extends the matching problem to more than two decks, with the possibilities of two-card matches, three-card matches, etc. Here, only his results for two decks are used.

If the two decks have the composition specified in (7) above, then the mean number of matches, ν, and the variance, σ^2, are given by:

$$\nu = \frac{1}{n} \sum m_i n_i$$

$$\sigma^2 = \frac{1}{n^2(n-1)} [(\sum m_i n_i)^2 - n \sum m_i n_i (m_i + n_i) + n^2 \sum m_i n_i]. \tag{10}$$

These formulas have been rearranged slightly from the form in which Battin listed them, for convenience in calculating. The products $m_i n_i$ and $m_i n_i (m_i + n_i)$ can be obtained quickly from the table (7) giving the composition of the deck, and the rest of the calculation is straightforward. We shall outline the calculations for the handwriting example.

i	1	2	3	Total
m_i	5	2	3	10
n_i	2	4	4	10
$m_i n_i$	10	8	12	30
$m_i + n_i$	7	6	7	
$m_i n_i (m_i + n_i)$	70	48	84	202

$$\nu = 30/10 = 3$$

$$\sigma^2 = [1/(100)(9)][(30)^2 - 10(202) + 100(30)]$$

$$\sigma^2 = 1880/900 = 2.0889$$

$$\sigma = \sqrt{2.0889} = 1.45.$$

Since the normal is a continuous distribution, while the matching problem distribution is discrete with probability "concentrated" on the integers, we must make a correction when using the normal as an approximation. That is, instead of the probability that $X = h$, we want the probability that X is between $h - 1/2$ and $h + 1/2$; if we want the probability of h or more matches, we must find $\text{Prob}(X \geq h - 1/2)$, where X is normally distributed with mean 3 and standard deviation 1.45. (In the special case where each deck has only two suits, the probability is concentrated on the even integers, and we must subtract 1 instead of 1/2 in order to get the proper correction.) Since we have available a table of the distribution of Y, a normal variate with zero mean and unit standard deviation, we look up

$$\text{Prob}\left(Y \geq \frac{h - \frac{1}{2} - 3}{1.45}\right).$$

The values found for $h = 0, \cdots, 7$ are listed in Table 3.

The last entry in Table 3 gives the binomial approximation to our matching problem. The binomial which came closest to fitting was that for which $n = 10$ and $p = .3$, which has mean 3 and variance 2.10.

REFERENCES

1. Anderson, T. W. On card matching. *Ann. math. Statist.*, 1943, **14**, 426-435.
2. Battin, I. L. On the problem of multiple matching. *Ann. math. Statist.*, 1942, **13**, 294-305.
3. Feller, W. An introduction to probability theory and its applications, Volume 1. New York: Wiley, 1950. pp. 62-69.
4. Greenwood, R. E. Probabilities of certain solitaire card games. *J. Amer. statist. Ass.*, 1953, **48**, 88-93.
5. Greville, T. N. E. Exact probabilities for the matching hypothesis. *J. Parapsychology*, 1938, **2**, 55-59.
6. Greville, T. N. E. The frequency distribution of a general matching problem. *Ann. math. Statist.*, 1941, **12**, 350-354.
7. Hamilton, C. H. Comment on question 8: variance of number of matches. *Amer. Statistician*, June, 1948, **2**, 23.
8. Huntington, E. V. Exact probabilities in certain card matching problems. *Science*, 1937, **86**, 499-500.
9. Olds, E. G. A moment-generating function which is useful in solving certain matching problems. *Bull. Amer. math. Soc.*, 1938, **44**, 407-413.
10. Rietz, H. L. Mathematical statistics, pp. 60-76 and 156-172. Chicago: Open Court Publishing Co., 1927.
11. U. S. Department of Commerce, National Bureau of Standards. Tables of the binomial probability distribution. Washington: U.S. Government Printing Office, 1949.

Manuscript received 1/10/55

Revised manuscript received 8/15/55

PSYCHOMETRIKA—VOL. 21, NO. 3
SEPTEMBER, 1956

A COMPARISON OF GAME THEORY AND LEARNING THEORY

HERBERT A. SIMON
CARNEGIE INSTITUTE OF TECHNOLOGY

It is shown that Estes' formula for the asymptotic behavior of a subject under conditions of partial reinforcement can be derived from the assumption that the subject is behaving rationally in a certain game-theoretic sense and attempting to minimax his regret. This result illustrates the need for specifying the frame of reference or set of the subject when using the assumption of rationality to predict his behavior.

Learning theory and game theory (together with the closely related statistical decision theory) purport to provide theories of *rational* behavior. Implicit in any theory of learning is a motivational assumption that learning consists in the acquisition of a pattern of behavior appropriate to *goal achievement, need reduction,* or the like. In parallel fashion game theory and statistical decision theory are concerned with discovering the course of action in a particular situation that will optimize the attainment of some objective pay-off.

In order to gain a better understanding of the concepts of *rationality* underlying these two bodies of theory, it would be interesting to construct a situation in which predictions made from these theories could be compared and then checked against experimental data on actual behavior. One situation of this kind received considerable attention at the Santa Monica Conference on Decision Processes (**2**, **3**, **4**). The experiment is one involving partial reinforcement. At each trial the subject chooses between two alternatives. Each alternative is rewarded on a certain per cent of the trials in which it is chosen (the trials rewarded being randomly determined) and not rewarded on the remaining trials in which it is chosen; the per cent of rewarded trials is in general different for the two alternatives. The learning theory advanced by Estes provides a prediction as to the frequency (in the limit as the number of trials increases) with which the first alternative will be chosen in preference to the second (**2**). The same frequency is predicted by the Bush-Mosteller theory when certain assumptions of symmetry are made with respect to the parameters that appear in their model (**1**, ch. 8). Estes reports several experiments that confirm predictions from his theory.

When this experimental situation was described to a number of game theorists at the Santa Monica conference, they pointed out that a *rational* individual would first estimate, by experimenting, which of the two alternatives had the greatest probability of reward, and would subsequently always

select that alternative which would not be predicted by the Estes theory. Flood has defended the choices predicted by the Estes theory against the charge of irrationality, basing his defense on two points (**4**, p. 288):

> (a) The proper definition of payoff utilities would be unclear in attempts to apply game-theoretic arguments to a real case, and there is a reasonable payoff matrix that would rationalize the reported behavior. Thus, if the organism's object were to maximize its score rather than its expectation, then it should sometimes not tend to use a pure strategy . . .
>
> (b) The von Neumann-Morgenstern game theory is inapplicable in this situation unless the organism can assume safely that the experimental stimulus is generated by a stationary stochastic process. For example, if the organism believes that there may be some pattern (non-stationarity) over time, in the stimulus, then it can often do better by using a mixed strategy rather than a pure one, for the latter would give it no way to discover any pattern effect.

In the next section, by combining in an appropriate fashion the two considerations advanced by Flood—that is, by assuming (a) the subject is trying to maximize something other than expected payoff, and (b) the subject does not believe or expect that the probability of reward from each alternative is fixed—it will be shown that the behavior predicted by the Estes theory and actually observed in experiments is rational in the sense of game theory (or at least in one of the many senses consistent with game theory). In a final section, the implications of this result will be discussed briefly.

Game-Theoretical Derivation of Estes' Result

Consider a partial reinforcement experiment in which there are two alternatives of behavior, A_1 and A_2 . If A_1 is chosen on a particular trial, it is rewarded with probability π_1 ; if A_2 is chosen, it is rewarded with probability π_2 . Let $p_1(t)$ be the probability that the subject chooses A_1 on the tth trial; $p_2(t) = (1 - p_1)$ the probability that he chooses A_2 . From the postulates of his learning model, Estes (**2**) predicts that the asymptotic value of p_1 as the number of trials increases will be p_1^* ,

$$p_1^* = \frac{1 - \pi_2}{(1 - \pi_1) + (1 - \pi_2)}. \tag{1}$$

This value for p_1^* may be obtained as the steady state of the stochastic process

$$\bar{p}(t + 1) = \Pi\bar{p}(t), \quad \text{where} \quad \Pi = \begin{pmatrix} \pi_1 & (1 - \pi_2) \\ (1 - \pi_1) & \pi_2 \end{pmatrix}; \tag{2}$$

for

$$p_1(t + 1) = \pi_1 p_1(t) + (1 - \pi_2)[1 - p_1(t)], \tag{3}$$

so that, if $p_1(t+1) = p_1(t) = p_1^*$,

$$(1 - \pi_1 + 1 - \pi_2)p_1^* = (1 - \pi_2), \tag{4}$$

from which (1) follows immediately.

We see that in Estes' theory π_1 is the probability that A_1 will be rewarded; but it is also the asymptotic probability that, having chosen A_1 on a given trial, the subject will choose it again on the next succeeding trial. A similar interpretation can be given to π_2 . Hence, we may interpret π_1 and π_2 as the conditional probabilities of *persistent* behavior when the subject has just chosen A_1 or A_2 , respectively; while $(1 - \pi_1)$ and $(1 - \pi_2)$ are the corresponding conditional probabilities of a *shift* in behavior.

For specificity, let us consider the case where $\pi_1 > \pi_2$. Then the game-theoretical objection to regarding as rational the asymptotic behavior predicted by the Estes model is that the subject could increase his expected reward by always choosing A_1 . For then the expected reward would be

$$\pi_1 > p_1^*\pi_1 + (1 - p_1^*)\pi_2 , \tag{5}$$

where the terms on the right-hand side of the inequality are easily seen to be the expected reward for the Estes model.

But the rationality of this game-theoretical solution is compelling only under the assumption that the reward probabilities are known to the subject, and known to be constant. These are the assumptions that Flood challenges. Let us consider an alternative set of assumptions which, while not the only possible such set, has some plausibility.

(*i*) The subject takes as given and fixed the π corresponding to the alternative he has chosen on the last trial. That is, he assumes the probability of reward to be π_1 or π_2 , if he persists in choosing again A_1 or A_2 , as the case may be.

(*ii*) The subject expects that if he *shifts* from the alternative just chosen to the other one, the probability of reward is unknown and dependent on a strategy of nature.

(*iii*) The subject does not wish to persist in his present behavior if there is a good chance of increased reward from shifting. He measures his success on each trial not from the reward received, but from the difference between the reward actually received and the reward that *could* have been attained if he had outguessed nature. In the terminology of L. J. Savage, he wishes to minimize his *regret*.

We may formalize these assumptions as follows: On each trial, the subject chooses between (*i*) *persisting* or (*ii*) *shifting* his choice. If he persists, he is rewarded with probability π (where $\pi = \pi_1$, or $\pi = \pi_2$ depending on whether the previous choice was A_1 or A_2 , respectively), irrespective of the strategy adopted by nature. If he shifts, he will receive a reward of 0 if nature adopts her strategy (α), and a reward of 1 if nature adopts her strategy

(β). The payoff matrix corresponding to these assumptions is:

	(α)	(β)
(i)	π	π
(ii)	0	1

where rows correspond to the subject's strategies and columns to nature's strategies. *Regret* is defined as the difference between the actual payoff for a given pair of strategies [e.g. (i), (β)], and the payoff that *could* have been realized, if the strategy actually employed by nature had been anticipated [e.g., (ii), (β)]. Performing the indicated subtractions, the regret matrix is:

	(α)	(β)
(i)	0	$(\pi - 1)$
(ii)	$-\pi$	0

(This was obtained from the first matrix by subtracting from each element the largest element in the same column).

Now let ρ be the probability that the subject uses strategy (i), i.e., persists, μ be the probability that nature uses strategy (α). Then the expected regret will be

$$\begin{aligned} R &= \rho\mu\cdot 0 + \rho(1-\mu)(\pi-1) + (1-\rho)\mu(-\pi) + (1-\rho)(1-\mu)\cdot 0 \\ &= \rho(1-\mu)(\pi-1) - (1-\rho)\mu\pi. \end{aligned} \tag{6}$$

The conditions that the regret be minimum (strictly, *minimax*) are given by

$$\frac{\partial R}{\partial \rho} = \frac{\partial R}{\partial \mu} = 0. \tag{7}$$

Using the second of these equations, we obtain from (6)

$$0 = -\rho(\pi - 1) - (1 - \rho)\pi, \tag{8}$$

whence

$$\rho = \pi. \tag{9}$$

Hence the subject would persist with probability π and shift with probability $(1 - \pi)$. But this is precisely the postulate contained in (2).

Hence, we have shown that the behavior predicted by Estes' theory is identical with that which would be exhibited by a *rational* subject intent on *minimaxing* his regret.

Comments on the Derivation

We need not try to decide whether the subjects who behave in conformity with the predictions of Estes' theory are minimaxing regret, or whether they are simply behaving in the adaptive fashion implied by the usual learning mechanisms. Most economists and statisticians would be tempted to accept the former interpretation, most psychologists the latter. It is not immediately obvious what source, other than introspection, would provide evidence for deciding the issue.

Perhaps the most useful lesson to be learned from the derivation is the necessity for careful distinctions between *subjective* rationality (i.e., behavior that is rational, given the perceptual and evaluational premises of the subject), and *objective* rationality (behavior that is rational as viewed by the experimenter). Because this distinction has seldom been made explicitly by economists and statisticians in their formulations of the problem of rational choice, considerable caution must be exercised in employing those formulations in the explanation of observed behavior.

To the experimenter who knows that the rewards attached to the two behaviors A_1 and A_2 are random, with constant probabilities, it appears unreasonable that the subject should not learn to behave in such a way as to maximize this expected gain—always to choose A_1 . To the subject, who perceives the situation as one in which the probabilities may change, and who is more intent on outwitting the experimenter (or nature) than on maximizing expected gain, rationality is something quite different. If rationality is to have any meaning independent of the perceptions of the subject we must distinguish between the rationality of the perceptions themselves (i.e., whether or not the situation as perceived is the real situation) and the rationality of the choice, given the perceptions.

If we accept the proposition that organismic behavior may be subjectively rational but is unlikely to be objectively rational in a complex world then the postulate of rationality loses much of its power for predicting behavior. To predict how economic man will behave we need to know not only that he is rational, but also how he perceives the world—what alternatives he sees, and what consequences he attaches to them (**5**). We should not jump to the conclusion, however, that we can therefore get along without the concept of rationality. While the Estes model predicts the behavior of naive subjects under partial reinforcement, we observe (**3**) that persons trained in game theory and placed in the same situation generally learn to choose A_1 consistently. It appears simpler to postulate here a change in set—a change in the perceptual model—rather than to attempt to explain this

behavior in terms of simpler learning-theoretic models. If anything was learned during the series of trials by the subjects who were game theorists, it was the appropriate perceptual model and not the appropriate behavior once that model is assumed.

REFERENCES

1. Thrall, R. M., Coombs, C. H., and Davis, R. L. (Eds.) Decision processes. New York: Wiley, 1954.
2. Estes, W. K. Individual behavior in uncertain situations: an interpretation in terms of statistical association theory. In (1), chap. 9.
3. Flood, M. W. Game-learning theory and some decision-making experiments. In (1), chap. 10.
4. Flood, M. W. Environmental non-stationarity in a sequential decision-making experiment. In (1), chap. 18.
5. Simon, H. A. A behavorial model of rational choice. *Quart. J. Econ.*, **69**, 1955, 99–118.

Manuscript received 4/18/55

Revised manuscript received 5/28/55

PSYCHOMETRIKA—VOL. 21, NO. 3
SEPTEMBER, 1956

"BEST POSSIBLE" SYSTEMATIC ESTIMATES OF COMMUNALITIES*

LOUIS GUTTMAN

THE ISRAEL INSTITUTE OF APPLIED SOCIAL RESEARCH

At least four approaches have been used to estimate communalities that will leave an observed correlation matrix R Gramian and with minimum rank. It has long been known that the square of the observed multiple-correlation coefficient is a lower bound to any communality of a variable of R. This lower bound actually provides a "best possible" estimate in several senses. Furthermore, under certain conditions basic to the Spearman-Thurstone common-factor theory, the bound must equal the communality in the limit as the number of observed variables increases. Otherwise, this type of theory cannot hold for R.

I. *Introduction*

One of the intriguing problems of factor analysis has been to find a formula for communalities that will minimize the rank of an arbitrary correlation matrix R. More explicitly, the problem is to find a diagonal matrix U such that $R - U^2$ is Gramian and of minimum rank.

Let n denote the order of R (and of U), and m the minimum rank for Gramian $R - U^2$. At least four approaches have been used to estimate communalities that will yield m:

(a) trial-and-error exact formulas
(b) exact formulas for special cases of R
(c) successive approximations
(d) lower bounds.

The main thesis of this paper is that, in certain senses, the last-mentioned of these four approaches provides "best possible" estimates of communalities for an arbitrary R, even though biased in general by being underestimates.

Let u_j be the jth diagonal element of any U that leaves $R - U^2$ Gramian (whether or not with minimum rank), and let h_j^2 be the corresponding communality:

$$h_j^2 = 1 - u_j^2 \qquad (j = 1, 2, \cdots, n). \tag{1}$$

Let ρ_j denote the multiple correlation coefficient of the jth variable in R on the remaining $n - 1$ variables, and σ_j the corresponding standard error

*This research was facilitated by a grant from the Lucius N. Littauer Foundation to the American Committee for Social Research in Israel in order to promote methodological work of the Israel Institute of Applied Social Research.

of estimate (assuming all observed variables to have unit variances):

$$\rho_j^2 = 1 - \sigma_j^2 \qquad (j = 1, 2, \cdots, n). \tag{2}$$

Then it has been shown (**2**, 92*f*; **3**, 293) that always

$$\rho_j^2 \leqq h_j^2 \qquad (j = 1, 2, \cdots, n). \tag{3}$$

No better general lower bound to h_j^2 has yet been established than ρ_j^2.

We shall prove here that there exist many nonsingular matrices R for which the equality in (3) holds for $n - m$ of the minimizing communalities—all but m of the ρ_j^2 are actual communalities (the remaining communalities equal unity). Such matrices R, however, are of a very restricted type.

A more generally useful result that we shall establish applies to the typical R postulated in the Spearman-Thurstone theory. This school of thought believes a common-factor analysis is meaningful only if m is small compared with n. We shall prove that if the ratio of m to n tends to zero as $n \to \infty$, then all except possibly a zero proportion of the ρ_j^2 must tend to the rank-minimizing h_j^2 . *If the Spearman-Thurstone hypothesis is correct for a given R, then the* ρ_j^2 *must almost always be very good approximations to the* h_j^2 *when n is large.* (Conversely, if the approximation is bad for many ρ_j , then the Spearman-Thurstone hypothesis of a limited number of common factors must be false.)

An even more general result refers to all R, regardless of the ratio of m to n. If there is to be one and only one unique-factor variable that can yield the uniqueness u_j^2 , then it must be that the limit of σ_j^2 must be u_j^2 as $n \to \infty$ (or it must be that $\rho_j^2 \to h_j^2$). Conversely, if σ_j^2 does not tend to u_j^2 as $n \to \infty$, then there is more than one "unique" variable that can provide the same loading u_j^2 (and satisfy all other algebraic requirements of common-factor theory); the larger the difference between σ_j^2 and u_j^2 (or between ρ_j^2 and h_j^2), the larger the possible difference between alternative "unique" parts for the same jth observed variable of R.

Other important properties of the lower bounds ρ_j^2 will be established. Before going on to our new results, it may be helpful to review briefly the four approaches listed above.

(*a*) *Trial and Error*

Assuming that sampling error and rounding-off errors in computations are nonexistent, trial and error is bound to yield an exact numerical answer when $m < n/2$; the diagonal elements of U^2 in such cases are *rational functions* of the non-diagonal elements of R (cf. **8**). It may turn out, of course, that U^2 is not uniquely determined; two or more different U^2 for the same R may yield m in many cases. When $m \geqq n/2$, trial and error can lead again to an expression for each communality, although in non-rational form in general. Again, multiple solutions for minimizing U^2 may occur.

(*b*) *Special Exact Formulas*

Some special cases of R make possible exact and rational formulas that need no apparent resort to trial and error. The known cases are for $m < n/2$, the most celebrated being Spearman's where $m = 1$. Thurstone has summarized a number of such formulas (**8**, ch. XIII). A caution should be added to Thurstone's discussion to the effect that not all the apparent solutions may yield Gramian U^2 nor leave $R - U^2$ Gramian. Actually these formulas beg the question, for it is generally not known in advance whether or not $m < n/2$. A specialized formula in effect must be tried on the given R to see if it works. Use of specialized formulas thus seems to be but a modified type of trial and error.

(*c*) *Successive Approximations*

Attempts have been made to avoid a direct exact solution for U^2 by taking recourse instead to successive approximations. An approximation U_1^2 is guessed, and $R - U_1^2$ is "factored" until residuals are considered small enough, leading to a second approximation U_2^2, etc. It has been claimed that such a procedure generally converges to a satisfactory U^2 (cf. **8**, p. 295). Algebraic proof of such convergence has never been published to our knowledge. For many iterative processes, the value to which convergence takes place depends on the initial trial value. That this may be the case for the above procedure seems evident when one recalls that there are many correlation matrices which do not have a unique set of communalities. Also, unless proof is given to the contrary, there is no reason to believe that successive approximations may not converge to some U where $R - U^2$ is *not* of minimum rank, if convergence takes place at all.

The issue of successive approximations is further beclouded by sampling considerations. Lawley's maximum likelihood solution seems the most appropriate put forward to date, as Rao points out (**7**). To attain precision in the sampling theory, apparently some restrictions have been introduced as to the nature of the population R, else the possibility of equally minimizing alternative solutions would remain. Again, it is not clear when a given R obeys these restrictions or when the sampling theory is valid in practice. [After the above was written, the writer received a copy of reference (**1**) in which a numerical example is given of the failure of Lawley's iterative procedure to converge properly.]

(*d*) *Lower Bounds*

If we again ignore sampling and rounding-off errors, it is always possible to establish useful lower bounds to communalities without any trial and error and without any hypothesis about or restrictions on R. The best of the lower bounds thus far established are the ρ_i^2, according to inequality (3)

above. It is often more convenient to discuss uniqueness rather than communalities, or to use inequality (4) rather than (3):

$$\sigma_j^2 \geqq u_j^2 \qquad (j = 1, 2, \cdots, n). \tag{4}$$

An important feature of the bounds in (3) and (4) is that they hold whether or not there is a multiple solution for U^2; they hold for all possible solutions simultaneously. Indeed, they lead to a criterion for choosing among multiple solutions, as indicated in the next section.

II. *Relationship to the Determinacy of Unique-Factor Scores*

Let r_j denote the multiple-correlation coefficient on the n observed variables of a unique-factor variable hypothesized to yield the uniqueness u_j^2. It has been shown in (6) that

$$r_j^2 = \frac{u_j^2}{\sigma_j^2} \qquad (j = 1, 2, \cdots, n). \tag{5}$$

Since (5) holds for all solutions U^2, it suggests that when a choice is necessary that which makes the inequalities (4) as small as possible is most desirable; the denominator on the right of (5) is fixed for j, so that such a choice makes the individual scores on the unique factor as determinate as possible from the observed data, or the r_j^2 as close as possible to unity. It has been shown that this also often tends to make individual scores on the *common*-factor variables as determinate as possible (6).

Should the approximations (4) for U^2 turn out *not* to be close in a given case, then the factor analysis itself may be regarded as not very useful or definitive. For it has been shown in (6) that determining factor *loadings* alone—common and unique—can be far from sufficient for pinning down *scores* on the hypothesized factors. *Alternative sets of scores for a given hypothesized factor can exist which yield identical loadings and yet correlate negligibly with each other,* according to formula (6),

$$r_j^* = 2r_j^2 - 1 \qquad (j = 1, 2, \cdots, n), \tag{6}$$

where r_j is given by (5) and r_j^* is the minimal correlation always attainable between two alternative sets of scores for the same unique factor hypothesized to underlie u_j^2. [According to (6), if $r_j^2 = .5$, then $r_j^* = 0$, or alternative score solutions for the same jth unique factor always exist that correlate *zero* with each other. Even if r_j^2 is as large as .9, this raises r_j^* only to .8. An equation parallel to (6) holds for common factors.]

III. *The "Best Possible" Estimates*

Can inequality (4) be improved on without recourse to some form of trial and error or use of specialized hypotheses? This does not seem possible. According to (5) this would imply some advance information on the r_j^2; there is no apparent way of getting such information on the r_j^2 in a

universally systematic manner. The situation seems to be the reverse: r_i^2 is determined by u_i^2 rather than conversely.

The rest of this paper will be devoted largely to showing that (4) is actually a "best possible" inequality in the sense that the phrase "best possible" is usually used mathematically for inequalities. The essential characteristics are that (a) many correlation matrices R exist for which the *equality* in (4) is actually attained at the same time that minimum rank m is attained, and (b) the inequalities in (4) must tend to equalities as n increases, under certain general conditions important to the theory of common-factor analysis. The bounds improve systematically in general as n increases, or as there is more information available from more observed variables. Furthermore, inequality (4) leads to inequalities for m that are also "best possible," and is closely related to the problem of estimating individual scores on the unique factors without any rank assumptions, via image analysis.

In virtually all attempts to solve the communality problem—whether exactly or by successive approximations—the problem is stated as for a fixed and finite n, or where R is from a finite number of n observed variables. It seems appropriate to ask also what happens to communalities as n increases or decreases.

While this issue is not discussed very explicitly by most writers, it usually seems implied that if the additional variables retain the same general kind of content as the initial ones, communalities of the initial ones should remain constant for all n sufficiently large. This would imply that for n small enough we should generally have $m > n/2$, or easy exact computations for U^2 (even ignoring sampling error) should be the exception rather than the rule. Having $m > n/2$ for relatively small n does not preclude m from remaining constant—and hence becoming relatively small—as n increases. It does imply that multiple solutions should be quite prevalent for finite n in practice. Furthermore, it cautions that an apparently exact solution for finite n may be but an artifact due to the finiteness of the number of variables observed.

It would be desirable, in view of all the preceding considerations, to have a systematic way of getting information about communalities with no assumptions whatsoever about R, yet without resorting to trial and error. Furthermore, this information should remain valid as n increases.

One of the virtues of the bounds (3) and (4) is that they possess these qualities in a simple and direct manner. This seems to be another type of "best possible" property from that usually considered, and one which appears peculiarly relevant to the problem of factor analysis.

IV. *Attaining Equality When n Is Finite*

If $\sigma_i^2 = 0$ for some j (so that $\rho_i^2 = 1$), then it must be that $u_i^2 = 0$ from (4) and the fact that a uniqueness cannot be negative. Here is one kind of

special circumstance wherein our bound becomes an exact estimate even when n is finite. In practice, this is not to be expected, since having one observed variable perfectly predictable from all the rest makes R singular.

Many cases of nonsingular R also exist for which the equality in (4) holds and n is finite. We shall exhibit some now. To this end, let us first recall that the σ_i^2 are the reciprocals of the corresponding main diagonals of R^{-1}. The following notation will be useful here and also later. Let S^{-2} (the inverse of S^2) be the diagonal matrix with the same main diagonal elements as R^{-1}. Then the jth main diagonal element of S^2 itself is simply σ_j^2 ($j = 1, 2, \cdots, n$):

$$S^2 = [\sigma_1^2, \sigma_2^2, \cdots, \sigma_n^2]. \tag{7}$$

If R is nonsingular, there exists a nonsingular matrix F such that

$$R = FF'. \tag{8}$$

F can be chosen in infinitely many ways for (8) when $n \geqq 2$, but always we can rearrange variables to find an F of the form

$$F = \begin{Vmatrix} A & 0 \\ B & C \end{Vmatrix}, \tag{9}$$

where A is a nonsingular square submatrix of order m, B is of order $(n - m) \times m$, and C is nonsingular and of order $n - m$. From (8) and (9),

$$R = \begin{Vmatrix} AA' & AB' \\ BA' & BB' + CC' \end{Vmatrix}. \tag{10}$$

It is easily verified that the inverse of F is given by

$$F^{-1} = \begin{Vmatrix} A^{-1} & 0 \\ -C^{-1}BA^{-1} & C^{-1} \end{Vmatrix}. \tag{11}$$

From (8), $R^{-1} = (F^{-1})'F^{-1}$, or using (11)

$$R^{-1} = \begin{Vmatrix} G & H' \\ H & (CC')^{-1} \end{Vmatrix}, \tag{12}$$

where

$$G = (AA')^{-1} + (A^{-1})'B'(CC')^{-1}BA^{-1} \tag{13}$$

and

$$H = -(CC')^{-1}BA^{-1}. \tag{14}$$

Now consider the special case where CC' is a diagonal matrix. Then $(CC')^{-1}$ is also diagonal. According to (12) and (7), $(CC')^{-1}$ constitutes the

lower right-hand submatrix of S^{-2}, or CC' constitutes the corresponding submatrix of S^2 and defines the σ_j^2 for $j = m + 1, m + 2, \cdots, n$. If we subtract this submatrix CC' from the lower right-hand corner of R in (10), we are clearly left with a reduced R that is Gramian and of rank m, it being the product of $[A\ B]'$ and its transpose. Thus we have:

Theorem 1. *If R can be factored into an F of the form* (9) *where CC' is diagonal* (*and A and C are nonsingular*), *then the main diagonal elements of CC' are the respective σ_j^2 for $j = m + 1, m + 2, \cdots, n$. If these $n - m$ σ_j^2 in CC' are subtracted from the corresponding main diagonal elements of R, the resulting matrix will be Gramian and of rank m.*

According to Theorem 1, when m is the actual minimal rank possible for Gramian $R - U^2$, then the first m diagonal elements of U^2 can be set equal to zero, and the last $n - m$ diagonal elements equal to the corresponding σ_j^2 as given by CC'. Thus, the last $n - m$ of the σ_j^2 serve exactly as rank-minimizing uniquenesses, or the equality in (4) holds for $j = m + 1, m + 2, \cdots, n$.

Notice that the *first* m uniquenesses implied by Theorem 1 are zero and not equal to the σ_j^2. If the first m σ_j^2 were also subtracted out from the main diagonal of R, then the resulting $R - S^2$ would in general not be Gramian, nor of minimum rank (cf. **4**).

Theorem 1 holds even when the m in it is not minimal. It is always possible to use the Theorem for the case where C is of order 1, and hence CC' is *necessarily* a diagonal matrix. This provides:

Corollary. For any nonsingular R, if any one σ_j^2 is subtracted from the corresponding main diagonal element of R, then the resulting matrix is of rank $n - 1$.

This result was partly indicated by Thurstone in his discussion of the "diagonal" method of matrix factoring (better known to mathematicians as the Schmidt or Gram-Schmidt process of orthogonalization), but without noticing apparently that his implied uniqueness was exactly σ_n^2 (**8**, p. 308).

We have thus completed showing that there are many matrices for which many of the σ_j^2 can serve as rank-minimizing uniquenesses. Also, we have the curious result that any one of the σ_j^2 alone will reduce nonsingular R to a Gramian matrix of rank $n - 1$.

V. *Equality in the Limit as $n \to \infty$*

We have already seen in Part III how, if a "unique"-factor variable is really to be uniquely determined for a given u_i^2, then we must have $u_i^2/\sigma_i^2 \to 1$ as $n \to \infty$, according to (5) and (6). This conclusion does not depend on the size of m, nor in particular on whether m remains finite or becomes infinite as $n \to \infty$. It thus applies to ordered factor theories—such as the radex,

with its simplexes and circumplexes (**5**)—as well as to limited common-factor theories like those of Spearman and Thurstone, whenever the δ-law of deviation (**5**, p. 308) holds for the unique-factor variables.

Thus, a general sufficient condition for σ_i^2 to tend to u_i^2 when $u_i^2 > 0$ is that $r_i^2 \to 1$ or $r^* \to 1$ as $n \to \infty$. This holds for each j separately.

A less general sufficient condition, and one that does not necessarily hold for any one j but only for "almost all" j, is given in

Theorem 2. If R is nonsingular for all n, and if $\lim_{n\to\infty} m/n = 0$, then

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \frac{u_i^2}{\sigma_i^2} = 1.$$

For all except possibly a zero proportion of the j it must be that $\lim_{n\to\infty} u_i^2/\sigma_i^2 = 1$.

The condition that $m/n \to 0$ holds in particular for the Spearman-Thurstone approach to factor analysis, which postulates that the number of common factors should be small compared to the number of observed variables.

Since $u_i^2/\sigma_i^2 \leqq 1$ for all j, according to (4), we must have the mean ratio also bounded above by unity:

$$\frac{1}{n} \sum_{i=1}^{n} \frac{u_i^2}{\sigma_i^2} \leqq 1. \tag{15}$$

The hypothesis that R is nonsingular for all n ensures that no observed variable is perfectly predictable from the rest, or that $\sigma_i^2 > 0$ for all j and n, so that division by σ_i^2 in (15) is always justified. The first conclusion of Theorem 2 is that the limit of the left member of (15) as $n \to \infty$ is actually the right member. But clearly, the mean value of a sequence cannot tend to an upper bound to each member of the sequence unless almost all members of the sequence also tend to this upper bound. Hence the second conclusion of Theorem 2 follows from the first. We need only to establish the first part of the theorem now.

As is well known, if $R - U^2$ is Gramian and of rank m, we can write

$$R = AA' + U^2, \tag{16}$$

where A is some matrix of order $n \times m$ and of rank m. Let Q be defined as the symmetric matrix of order m:

$$Q = I_m + A'U^{-2}A, \tag{17}$$

where I_m is the unit matrix of order m. It has been shown in (**2**, 92) that Q is Gramian and nonsingular, and furthermore

$$R^{-1} = U^{-2} - U^{-2}AQ^{-1}A'U^{-2}. \tag{18}$$

It is easily verified further, from (18) and (17), that

$$A'R^{-1}A = I_m - Q^{-1}. \tag{19}$$

Since the left member of (19) is clearly Gramian, so must the right member be. Indeed, it is known that $I_m - Q^{-1}$ is the covariance matrix of the predicted values (from the observed n variables) of any m orthogonal common-factor scores underlying loading matrix A (6). Let q^{kk} denote the kth main diagonal element of Q^{-1}, or the variance of estimate of the kth common factor, and let p_k^2 be defined as

$$p_k^2 = 1 - q^{kk} \qquad (k = 1, 2, \cdots, m). \tag{20}$$

Then p_k^2 is the square of the multiple-correlation coefficient of the kth common factor from the n observed variables, and

$$0 \leqq p_k^2 \leqq 1 \qquad (k = 1, 2, \cdots, m). \tag{21}$$

Therefore, the *trace*—or sum of the main diagonal elements—of $I_m - Q^{-1}$ satisfies

$$\text{tr}\,(I_m - Q^{-1}) = \sum_{k=1}^{m} p_k^2 \leqq m. \tag{22}$$

We are particularly interested in the trace of U^2R^{-1}, for clearly—remembering (7)—

$$\text{tr}\,(U^2R^{-1}) = \text{tr}\,(U^2S^{-2}) = \sum_{i=1}^{n} \frac{u_i^2}{\sigma_i^2}. \tag{23}$$

Since the trace of a product is unchanged if order of multiplication is reversed,

$$\text{tr}\,(A'R^{-1}A) = \text{tr}\,(AA'R^{-1}) = \text{tr}\,(I_n - U^2R^{-1}), \tag{24}$$

the last member following from the middle member by recalling (16). Therefore, taking traces of both members of (19) and using (23), (24), and (22),

$$\sum_{i=1}^{n} \frac{u_i^2}{\sigma_i^2} = n - \sum_{k=1}^{m} p_k^2 \geqq n - m. \tag{25}$$

Dividing (25) through by n and prefixing inequality (15),

$$1 \geqq \frac{1}{n} \sum_{i=1}^{n} \frac{u_i^2}{\sigma_i^2} = 1 - \frac{1}{n} \sum_{k=1}^{m} p_k^2 \geqq 1 - \frac{m}{n}. \tag{26}$$

Clearly, if $m/n \to 0$ in the last member of (26), the middle members must tend to unity, or Theorem 2 is established.

Notice that Theorem 2 could be rephrased to say that almost all $r_i^2 \to 1$, or almost all unique-factors must be determinate in the limit. It is interesting to see this in a slightly different way. From (5) and the middle members of (26),

$$\frac{1}{n} \sum_{i=1}^{n} r_i^2 + \frac{1}{n} \sum_{k=1}^{m} p_k^2 = 1, \tag{27}$$

or

$$\overline{r^2} + \frac{m}{n}\overline{p^2} = 1, \tag{28}$$

where $\overline{r^2}$ is the mean predictability of the n unique-factors, while $\overline{p^2}$ is the mean predictability of the m common factors. When m/n is small, the average predictability of the common factors cannot influence greatly the average predictability of the unique factors: $\overline{r^2}$ must be close to unity. A further consequence is that, if both $\overline{r^2}$ and $\overline{p^2}$ tend to unity as $n \to \infty$, it must be that $m/n \to 0$. This does not require m to remain finite, of course, but only to increase at a less rapid pace than does n.

VI. *Increase in Information with n*

A desirable property of estimates of communalities is that they should improve in general as n increases. Any n variables studied empirically by a factor analysis are usually regarded as but a sample of a far larger universe of variables. The communalities sought are those of the universe.

Of the four approaches to estimates outlined in Part I above, the only one which has its estimates vary explicitly with n is that of lower bounds. In this sense, it is the only one not tied to algebraic artifacts that may arise in data due to the finiteness of n of the observed sample of variables (cf. **3** and **4**).

For fixed j, ρ_j^2 must increase with n—or at worst remain constant—for a multiple-correlation cannot become worse as the number of predictors increases. If h_j^2 is defined as for the universe of variables ($n = \infty$), then ρ_j^2 must improve in general as an estimate of h_j^2 as n increases, considering (3). The lower bounds improve as estimates as n increases, taking advantage of the increased information.

Similarly, if the jth unique-factor scores are defined uniquely as for the universe of observed variables, r_j^2 must in general increase with n. From (5), this again makes σ_j^2 an increasingly better estimate of the fixed u_j^2 as n increases.

Thus, the lower bounds automatically take advantage of whatever new information is brought in with increased n, without making any assumptions at all. In broad classes of cases, as we have seen, this new information can make $\rho_j^2 \to h_j^2$ for all or almost all j.

VII. *Further "Best Possible" Inequalities*

We have concentrated until now on the approximation of the σ_j^2 to the u_j^2. Related to this is another problem: the estimation of minimum rank m for Gramian $R - U^2$. We shall show that using the diagonal matrix S^2 of (7) as an estimate of U^2 for finite n leads also to a "best possible" inequality for m, as well as to other important inequalities.

With any nonsingular correlation matrix R is associated another nonsingular correlation matrix R^* defined by

$$R^* = SR^{-1}S. \tag{29}$$

R^* is clearly Gramian, for R^{-1} is Gramian and S is a diagonal matrix. The main diagonal elements of R^* are all unity from the definition of S and the fact that $1/\sigma_j^2$ is the jth diagonal element of R^{-1}. Indeed, R^* is the correlation matrix of the *anti-images* of the n variables of R (cf. **3**, p. 294f). Regardless of the statistical meaning of R^*, it is a perfectly good correlation matrix when n is finite, and we can seek a diagonal Gramian matrix U^* that will leave $R^* - U^{*2}$ Gramian and with minimum rank m^*. This will lead to the interesting and important inequality for the case where no σ_j^2 is a uniqueness nor equals unity:

$$m + m^* \geqq n \qquad (u_j^2 < \sigma_j^2 < 1; j = 1, 2, \cdots, n). \tag{30}$$

The restrictions that $S^2 - U^2$ and $I - S^2$ be nonsingular are essential here (consider the counter-example where $S = R = R^* = I$). That $\sigma_j^2 \neq 1$ ($I - S^2$ be nonsingular) implies that each variable in R has at least one nonzero correlation with some other variable.

According to (30), if m/n is small, then m^*/n must be large. Conversely, if m^*/n is small, m/n must be large. This is rather paradoxical in view of the fact that R^* can *always* be reduced to rank m by subtracting out the diagonal matrix $SU^{-2}S$ $(= S^2U^{-2})$. This follows by pre- and post-multiplying (18) through by S, remembering (29), and noting that the second term on the right is of rank m. Conversely, R can always be reduced to rank m^* by subtracting out $S^{*2}U^{*-2}$, where S^{*2} is the diagonal matrix defined by the main diagonal of R^{*-1}. Thus, if all diagonal-free submatrices of R have rank less than $n/2$, so must those of R^*, and conversely. Regardless, (30) holds.

In effect, then, (30) implies that to every R for which $\sigma_j^2 \neq u_j^2$ or 1 for all j and where $m < n/2$, there corresponds an R^* which is a generalized "Heywood" case (cf. **4**, 159f). Although all diagonal-free matrices have small rank in R^*, no communalities can be found to make $R - U^2$ of equally small rank and yet be Gramian. It must be that $m^* \geqq n - m$. This again emphasizes that the case $m < n/2$ may be the exception, rather than the rule, for correlation matrices. And it is interesting that this paradox arises precisely for those cases where no σ_j^2 equals the corresponding u_j^2.

To establish (30), we first recall the theorem (**4**, 157f) that if $S^2 - U^2$ is nonsingular, and if s is the non-negative index of $R - S^2$, then

$$s \leqq m \qquad (|\, S^2 - U^2 \,| > 0). \tag{31}$$

Now, the proof of (31) in (**4**) can be modified to take care of the case where $S^2 - U^2$ is possibly singular, to establish the weaker but more universal inequality $p \leqq m$, where p is the *positive* index of $R - S^2$. We shall not take

space to prove this modification here, but shall merely state it in terms of our needs for R^*:

$$p^* \leqq m^*, \tag{32}$$

where p^* is the positive index of $R^* - S^{*2}$, and we do *not* necessarily assume $S^{*2} - U^{*2}$ to be nonsingular.

Now, from (29), $R^{*-1} = S^{-1}RS^{-1}$, or since the main diagonal elements of R are all unity,

$$S^{*2} = S^2. \tag{33}$$

It is interesting to note that (33) and (29) imply that $(R^*)^* = R$, or R is to R^* as R^* is to R.

Statistically, (33) implies that the relative predictability of the jth anti-image from the $n - 1$ remaining anti-images is the same as for the jth original variable from the $n - 1$ remaining original variables. From (29) and (33) we can write the identity

$$R^* - S^{*2} = S(R^{-1} - I)S. \tag{34}$$

Sylvester's "law of inertia" (cf. **4**, p. 152) applied to (34) shows that p^* equals the positive index of $R^{-1} - I$, which in turn clearly equals the number of latent roots of R^{-1} greater than unity. Hence p^* equals the number of latent roots of R itself which are *less* than unity. But it has been shown in (**4**) that s is not less than the number of latent roots of R which are greater than or equal to unity whenever $I - S^2$ is nonsingular. Since R has n latent roots all told, it follows that

$$s + p^* \geqq n \qquad (|\, I - S^2 \,| > 0). \tag{35}$$

Inequality (30) follows from (31), (32), and (35).

To prove that (30) is a "best possible" inequality, we must show that matrices R exist for which the equality sign holds. It suffices to consider an R which has only two distinct latent roots, say $\lambda_1 > 1$ with multiplicity f and $\lambda_2 < 1$ with multiplicity $f^* = n - f$. Then it must be that

$$m = f, \qquad m^* = f^*. \tag{36}$$

For $m \geqq f$ by inequality (39) of (**4**, 159), and hence $m = f$ by considering that $R - \lambda_2 I$ is Gramian and of rank f; $m^* = f^*$ by analogous reasoning on R^{-1}. Since $f + f^* = n$, (36) provides a special case where the equality in (30) holds.

Inequality (31) by itself is similarly a "best possible" one. Consider the case where R^* has two distinct latent roots, say $\lambda_1 < 1$ with multiplicity p^* and $\lambda_2 < 1$ with multiplicity $p = n - p^*$. Since $R^{*-1} - I = S^{-1}RS^{-1} - I = S^{-1}(R - S^2)S^{-1}$, p is the positive index of $R - S^2$ while p^* is that of $R^* - S^{*2}$. Also, since no root vanishes, $p = s$ or the positive and non-negative

indices coincide. Since $R^{*-1} - \lambda_2^{-1}I = S^{-1}RS^{-1} - \lambda_2^{-1}I$ is Gramian and of rank $p = s$, so must $R - \lambda_2^{-1}S^2$ be, or the equality in (31) must hold for this case.

VIII. *Relation to Image Analysis*

The ratio of u_i^2 to σ_i^2 indicates the relative predictability of the jth unique-factor scores from the n observed variables of R, according to (5). Closely related is another parameter developed in image theory and denoted by δ_i^2, namely, the variance of the difference between the respective scores on the jth anti-image and the jth unique factor. It turns out (**3**, 293) that δ_i^2 can be computed as the simple difference

$$\delta_i^2 = \sigma_i^2 - u_i^2 \qquad (j = 1, 2, \cdots, n). \tag{37}$$

Hence, a necessary and sufficient condition that $\sigma_i^2 \to u_i^2$ as $n \to \infty$ is that $\delta_i^2 \to 0$. This implies that the unique-factor scores must be essentially the total anti-image scores from the universe of content. Here we have the individual anti-images themselves as increasingly better estimates of the unique-factor scores as $n \to \infty$. This problem of estimating scores is perhaps even more basic than that of estimating only over-all parameters, such as uniquenesses, which are based on the scores. Estimating U^2 by S^2 has the important property of tying in directly with the score estimation problem via image analysis.

REFERENCES

1. Golub, G. H. On the number of significant factors as determined by the method of maximum likelihood. (Stencilled) Digital Computer Laboratory, Univ. of Illinois, 1954.
2. Guttman, L. Multiple rectilinear prediction and the resolution into components. *Psychometrika*, 1940, **5**, 75-99.
3. Guttman, L. Image theory for the structure of quantitative variates. *Psychometrika*, 1953, **18**, 277-296.
4. Guttman, L. Some necessary conditions for common-factor analysis. *Psychometrika*, 1954, **19**, 149-161.
5. Guttman, L. A new approach to factor analysis: the radex. In Paul F. Lazarsfeld (ed.), Mathematical thinking in the social sciences, Glencoe, Ill.: Free Press, 1954.
6. Guttman, L. The determinacy of factor score matrices, with implications for five other basic problems of common-factor theory. *Brit. J. stat. Psychol*, 1955, **8**, 65-81.
7. Rao, C. R. Estimation and tests of significance in factor analysis. *Psychometrika*, 1955, **20**, 93-111.
8. Thurstone, L. L. Multiple-factor analysis. Chicago: Univ. Chicago Press, 1947.

Manuscript received 5/17/55

Revised manuscript received 9/13/55

PSYCHOMETRIKA—VOL. 21, NO. 3
SEPTEMBER, 1956

RANK-BISERIAL CORRELATION

EDWARD E. CURETON

UNIVERSITY OF TENNESSEE

A formula is developed for the correlation between a ranking (possibly including ties) and a dichotomy, with limits which are always ± 1. This formula is shown to be equivalent both to Kendall's τ and Spearman's ρ.

Suppose we have two correlated variables, one represented by a ranking (possibly including ties) and the other by a dichotomy. The dichotomy may be considered a ranking concentrated into two multiple ties; its ties, however, do not represent equal measurements (or judgments of equality) on a continuous (or at least a many-step) variable. Rather, the ties represent a broad grouping of the data into two categories, or possibly an actual two-point distribution (sex, e.g.). Since the number of distinct ranks in the ranked variable will always be much greater than 2 and will equal N in the untied case, exact rank agreement of the two variables, pair by pair for each individual, is impossible. In this situation we desire a coefficient which will still have attainable limits ± 1 in all circumstances. It should be $+1$ when all ranks in the "higher" category of the dichotomy exceed all ranks in the "lower" category, and -1 when all ranks in the "lower" category exceed all ranks in the "higher" category. It should be strictly non-parametric, i.e., defined wholly in terms of inversions and agreements between pairs of rank-pairs, without use of such concepts as mean, variance, covariance, or regression. Finally, it should resemble the usual rank correlation coefficients in some reasonable sense.

Let R_x represent the dichotomy, with categories $R_x +$ and $R_x -$, and let R_y represent the ranked variable. Ties in R_y are to be handled by the mid-rank method. We then arrange the ranks R_y in as nearly as possible the natural order $(N, N-1, \cdots, 1)$, with rank N "high" and rank 1 "low," and allocate them to the categories $R_x +$ and $R_x -$ as in the following example:

$R_x +$	$R_x -$	*Inv.*	*Agr.*
9.5			4
9.5			4
8			4
6.5			3
	6.5	2	
4.5			3
4.5			3
	2.5		
	2.5		
	1		
$N_1 = 6$	$N_2 = 4$	$Q = 2$	$P = 21$

(1)

No two R_y ranks may be in the same row, but in case of a tie in R_y with one member falling under $R_x +$ and the other under $R_x -$, the relation between the row and column allocations is immaterial. Thus, in (1), the first 6.5 might as well have been allocated to $R_x -$ and the second to $R_x +$.

With this arrangement, there is an *inversion* at any given number under $R_x -$ for every smaller number under $R_x +$. Thus, at 6.5 in $R_x -$ we have two inversions, one for each of the values 4.5 under $R_x +$. There is also an *agreement* at any given number under $R_x +$ for every smaller number under $R_x -$. Let Q be the total number of inversions, and let P be the total number of agreements.

With this method of allocation to rows and columns, perfect positive correlation would require that all numbers under $R_x +$ should be larger than all numbers under $R_x -$, and in this case we should find that $Q = 0$ and $P = P_{\max}$. Perfect negative correlation would require that all numbers under $R_x +$ should be smaller than all numbers under $R_x -$, and in this case we should find that $P = 0$ and $Q = Q_{\max}$. Also, $P_{\max} = Q_{\max}$, since the two result merely from an interchange of the sets of numbers under $R_x +$ and $R_x -$. Our coefficient may therefore be of the form

$$r_{RB} = (P - Q)/P_{\max} . \tag{2}$$

It will be $+1$ if $Q = 0$ and $P = P_{\max}$, -1 if $P = 0$ and $Q = Q_{\max} = P_{\max}$, and 0 if $P = Q$.

To determine $P_{\max}$, we note first that in the situation in which the coefficient is $+1$, there will be N_2 agreements for every number under $R_x +$, or N_1N_2 in all. There is one case, however, so far passed over, in which $P_{\max}$ cannot be as great as N_1N_2. This case is illustrated in our example. If we set up explicitly the situation for $P = P_{\max}$ with these data, we have:

$R_x +$	$R_x -$	*Inv.*	*Agr.*
9.5			4
9.5			4
8			4
6.5			4
6.5			4
4.5			3
	4.5		
	2.5		
	2.5		
	1		
$N_1 = 6$	$N_2 = 4$	$Q = 0$	$P = 23$
			$N_1N_2 = 24$

(3)

One agreement is lost because the lowest rank under $R_x +$ is tied with the highest under $R_x -$. In other cases there might be a triple or multiple tie at the point of dichotomy. We shall term a tie at this point a *bracket tie.* For any bracket tie, the value of P_{max} will be reduced from N_1N_2 by unity for every *pair* of members of this tie one of which is under $R_x +$ and the other under $R_x -$, after R_y has been rearranged to be as nearly as possible in the natural order and allocation under $R_x +$ and $R_x -$ is made in such a manner as to preserve the original values of N_1 and N_2. If t_1 is the number under $R_x +$ participating in the bracket tie, and t_2 the number under $R_x -$, $P_{max} = N_1N_2 - t_1t_2$, and our formula becomes

$$r_{RB} = \frac{P - Q}{N_1N_2 - t_1t_2}. \tag{4}$$

Physically, it is not necessary to rearrange the original data in order to compute t_1t_2. We merely draw a horizontal line across columns $R_x +$ and $R_x -$ in (1), at a level which leaves N_1 cases above the line and N_2 below it. Since the original arrangement in (1) was with R_y in as nearly as possible the natural order, a bracket tie will then consist of any group of identical numbers, some immediately above and some immediately below this line. The number above is t_1 and the number below is t_2. For the example of (1), we find by (4):

$$r_{RB} = \frac{21 - 2}{(6)(4) - (1)(1)} = .826.$$

Clearly r_{RB} is a Kendall-type coefficient, since Q and P are the numbers of unweighted inversions and agreements, respectively (2). But it is also a Spearman-type coefficient. Durbin and Stuart (1) have shown that, in the untied case, Spearman's coefficient is given by $(U - V)/(U - V)_{max}$, where

V is the number of inversions and U the number of agreements, each weighted by the difference between the two ranks concerned. It is easily shown that the difference which supplies the weight may come from either R_y or R_x, and it is also easy to find $(U - V)_{max}$ for the cases corresponding to Kendall's ρ_a and ρ_b. The writer has not been able to prove in these cases that the values given by $(U - V)/(U - V)_{max}$ are necessarily equal in general to those given by the corresponding formulas based on Σd^2, but he has verified each of them on several sets of numerical data.

In the present case, we need merely note that all R_x values bracketed under $R_x +$ would have one mid-rank value, and all those bracketed under $R_x -$ another. If, then, we weight each inversion and agreement by the corresponding rank-difference in R_x, all weights will be equal (and equal to the difference between the two mid-rank values), and it follows at once that r_{RB} is a Spearman-type coefficient.

The hypothesis that r_{RB} differs only by chance from $\rho_{RB} = 0$ may be tested by the Mann-Whitney extension of the Wilcoxon test (**3**).

REFERENCES

1. Durbin, J. and Stuart, A. Inversions and rank correlation coefficients. *J. roy. statist. Soc.*, 1951, **13**, 303-9.
2. Kendall, M. G. Rank correlation methods. London, Charles Griffin and Co., 1948, Ch. 3.
3. Mann, H. B. and Whitney, D. R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. math. Statist.*, 1947, **18**, 50-60.

Manuscript received 11/14/55

PSYCHOMETRIKA—VOL. 21, NO. 3
SEPTEMBER, 1956

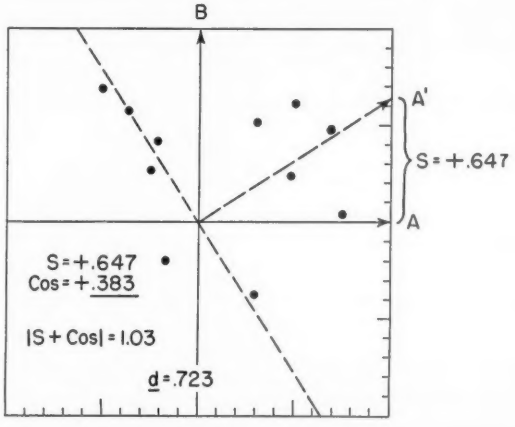
A NOMOGRAM FOR FACTOR ANALYSTS*

P. D. DEUEL

UNIVERSITY OF CALIFORNIA

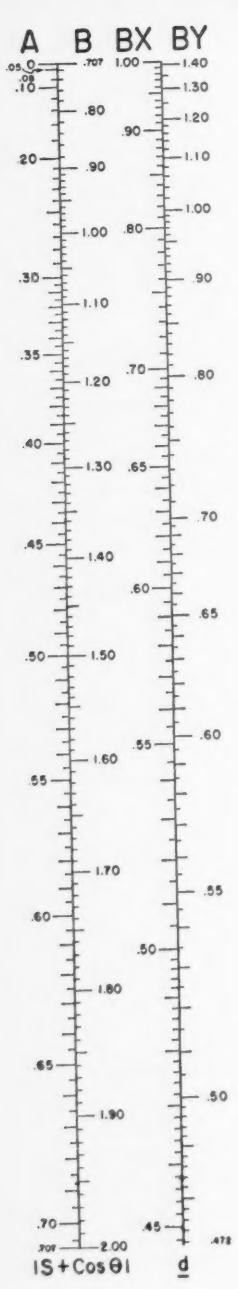
When a new reference vector is chosen graphically from the plane of two old ones, its direction cosines as well as the projections of the tests on it are most easily computed by applying certain multipliers d and Sd to quantities which are already known. The nomogram quickly supplies d, after S has been read from the graph.

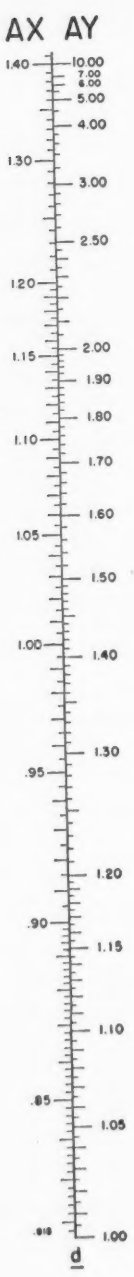
The nomogram accompanying this article reduces the computing work in what is perhaps the most popular of the graphical rotation methods of factor analysis, namely, the diagrammatic method explained in Thurstone (1, pp. 194–216) with rotations made in one plane at a time. Use of the nomogram will be explained in terms of the following figure:



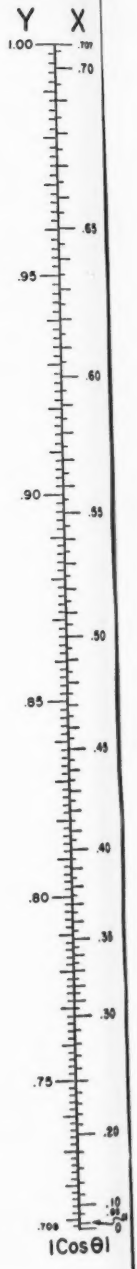
As usual, A and B represent trial reference vectors and the dots represent the projections of the test vector termini in the plane determined by A and B. (Following psychological usage, we refer to the "given" or analyzed variables as tests.) The cosine, .383, of the angle θ between A and B has been recorded in the lower left-hand portion of the diagram; it was obtained

*Suggestions by Norman Livson, Thomas Nichols, and Karyl Atherton have been incorporated in the nomogram. Mrs. Atherton also checked the necessary computations. In addition, I am obligated to Katherine Eardley, scientific illustrator, for her care in lettering and inking the original.





NOMOGRAM giving
$\underline{d} = (S^2+2S\cos\theta+1)^{-1/2}$
as a function of
$|S+\cos\theta|$ and $|\cos\theta|$
for $0 \leq |S+\cos\theta| \leq 2$



in verifying the linear independence of A and B. The figure suggests replacement of A by a new trial reference vector A'', collinear with the dashed-line vector A', for five tests have nearly vanishing projections on A'. From the diagram we read the vector equation $A' = A + .647B$. The projections of all the tests on A'', as well as the direction cosines of A'', are required.

More generally, the method under discussion will always yield a "long reference vector" A' related to known reference vectors A and B by one of the equations $A' = A + SB$ or $-A' = A + SB$, for some number S between -1 and 1. When $A' = A + SB$, the inner product (V, A'') between any vector V and the unit vector A'' collinear with A' is

$$(V, A'') = d(V, A) + S\,d(V, B), \tag{1}$$

where d is $(A', A')^{-\frac{1}{2}}$, the reciprocal of the length of A'. If V is a test vector, (1) gives the desired projection (V, A'') of V on A'' in terms of the known projections (V, A) and (V, B) of V on A and B. Similarly, if V is one of the orthogonal basic vectors, (1) gives the desired direction cosine (V, A'') of A'' with respect to V in terms of the given direction cosines (V, A) and (V, B) of A and B with respect to V. Thus, when d is known, all the required quantities are obtained from (1) by applying the multipliers d and Sd to columns containing the relevant inner products (V, A) and (V, B). The case $-A' = A + SB$ is accommodated by simply negating the right side of (1).

Upon noting that (A, B) is the cosine of the angle θ between A and B, it is seen that

$$d = (|\, S + \cos\theta\,|^2 - |\cos\theta|^2 + 1)^{-1/2}, \tag{2}$$

a graphable function of $|\, S + \cos\theta\,|$ and $|\cos\theta|$, and in fact, the function represented in the nomogram. To determine d, therefore, it is only necessary to perform the addition $S + \cos\theta$, enter the nomogram with arguments $|\, S + \cos\theta\,|$ and $|\cos\theta|$, and read d from the appropriate inside scale. To read the nomogram:

1. Locate $|\, S + \cos\theta\,|$ on left-most stem and $|\cos\theta|$ on right-most stem.
2. Align a straightedge through the two points thus found.
3. Pick out appropriate inside scale. This is the scale whose label (at top) combines the labels of the scales on which $|\, S + \cos\theta\,|$ and $|\cos\theta|$ were found.
4. Read d where straightedge crosses appropriate inside scale.

Generally, S is read from a graph and cannot be identified with better than two-place accuracy. In such cases a gratuitous third digit may be appended to S, so chosen as to render the third digit of the sum $S + \cos\theta$ zero. This device makes for greater ease and precision in locating values of $|\, S + \cos\theta\,|$ near the top of the left-hand stem.

In our example, the value $+.64 \pm$ for S was read from the graph, and the final digit 7 was selected to complement the third digit 3 of $\cos \theta$. The addition $S + \cos \theta = 1.03$ was performed directly on the diagram. The value 1.03 for $| S + \cos \theta |$ was then found on the B-scale of the nomogram; the value .383 for $| \cos \theta |$ was located on the X-scale; hence, $d = .723$ was read from the BX-scale.

The scale factor of the nomogram varies widely from the lower part of the BX-scale to the upper region of the AY-scale. The instrument has been designed, however, with the purpose of securing three-place accuracy for virtually all cases which arise in practice; extensive applications of the nomogram, both to Thurstone's illustrative material and to original factor analyses, indicate that this goal has been well attained. In successive rotations of the above type, the liberal AX-scale carries by far the most traffic.

Larger copies of the nomogram may be obtained by writing to the author at the Institute of Child Welfare, University of California, Berkeley 4, California. The copies have stems approximately $9\frac{1}{4}''$ long and are printed on $8\frac{1}{2}''$ x $11''$ index-card paper.

REFERENCE

1. Thurstone, L. L. Multiple-factor analysis. Chicago: Univ. Chicago Press, 1947.

Manuscript received 8/29/55

Revised manuscript received 10/24/55

PSYCHOMETRIKA—VOL. 21, NO. 3
SEPTEMBER, 1956

A NOTE ON THE ESTIMATION OF NONSPURIOUS CORRELATIONS

WILLIAM H. ANGOFF

EDUCATIONAL TESTING SERVICE

A method is provided for estimating the nonspurious correlation of a part of a test with the total test. Two cases are considered: one in which the actual subtest is parallel to the total test, the other in which the actual subtest is not parallel to the total test.

A problem that frequently arises in the examination of test data is that of estimating the degree of relationship of a test with a subsection of items drawn from the test. In general two methods are available: one is to correlate the subtest with the total parent test; the other is to correlate the subtest with the complementary subtest resulting from the subtraction of the first subtest from the total. In the first method, it is clear that spuriousness exists; even if the subtest is totally unreliable and correlates zero with the complementary subtest, the computed correlation of the part with the total would result in a value greater than zero, and roughly in the proportion represented by the subtest in the total. In general the correlation would be

$$r_{jt} = (\sigma_j + r_{jh}\sigma_h)/\sigma_t, \tag{1}$$

where t is the total test, and j and h are the complementary subsections of the total test. Even if $r_{jh} = 0$, r_{jt} is still greater than zero, merely by virtue of the presence of j in the total test.

The second method, on the other hand, defeats its own purpose—it yields a correlation of the subsection j with the complementary subsection h, not a correlation with a test of the length t.

In order to estimate the nonspurious correlation, say $r_{j't}$, consider test t to be an unspeeded test of power, and, as before, to comprise two parts, j and h. Also consider a hypothetical test, j', exactly parallel and of equivalent effective length (1) to j. Subtests j and h need not be parallel forms. Then

$$r_{j't} = (r_{jj'}\sigma_j + r_{j'h}\sigma_h)/\sigma_t . \tag{2}$$

The notation in (2) may be modified slightly. Since j and j' are parallel, $r_{j'h}$ may be written r_{jh}. Also $r_{jj'}$, which expresses the reliability of test j as the correlation between parallel forms, may be written in the conventional notation as r_{jj}. Henceforth, $r_{j'h}$ will be written as r_{jh}, and $r_{jj'}$ will be written as r_{jj}. However, the notation $r_{j't}$, designating nonspurious correlation, will

be retained to distinguish it from the spurious correlation, r_{jt} . With the foregoing modifications in notation the formula for nonspurious correlation is

$$r_{j't} = (r_{jj}\sigma_j + r_{jh}\sigma_h)/\sigma_t \ . \tag{3}$$

In the actual situation, the reliability, r_{jj} , will probably be estimated best by one of the internal consistency reliability formulas appropriate to power tests.

If now a further restriction is placed on test t, namely, that j and h be parallel (but not necessarily of equivalent length), then further simplification is possible. It has been observed (1) that under this restriction

$$r_{jj} = r_{tt}r_{jt}^2 \ , \tag{4}$$

and also

$$r_{tt} = \frac{r_{jt}\sigma_t - \sigma_j}{r_{jt}(\sigma_t - r_{jt}\sigma_j)}. \tag{5}$$

Substituting (5) in (4),

$$r_{jj} = \frac{r_{jt}(r_{jt}\sigma_t - \sigma_j)}{\sigma_t - r_{jt}\sigma_j}. \tag{6}$$

It may also be seen that if $r_{jt}\sigma_t - \sigma_j$ is substituted in (3) for its equivalent, $r_{jh}\sigma_h$ [see equation (1)], then

$$r_{j't} = (r_{jj}\sigma_j + r_{jt}\sigma_t - \sigma_j)/\sigma_t \ . \tag{7}$$

Finally, substituting (6) in (7),

$$r_{j't} = \frac{r_{jt}\sigma_t - \sigma_j}{\sigma_t - r_{jt}\sigma_j}. \tag{8}$$

Two formulas are thus presented for estimating the nonspurious correlation of a subtest of items with the total test from which it is drawn. In (3) and in its equivalent, (7), no restriction is imposed on the kinds of items drawn from the total test. These equations could, for example, represent the estimated correlation of a subset of arithmetic items drawn from a heterogeneous total test consisting of arithmetic, verbal, and spatial items. Equation (8), on the other hand, requires that the subset of items be essentially a short parallel form or miniature of the total test. Whereas (3) and (7) require a separate determination of the reliability of the subtest by means of internal consistency methods, such as the Kuder-Richardson formulas, (8) permits that estimate to be made implicitly as an integral part of the estimate of nonspurious correlation. It is the added restriction that j is parallel to t that makes the simplification possible.

Additional algebraic simplifications may be made in (8). From (1) it is seen that the numerator of (8) may be written $r_{jh}\sigma_h$. It may also be seen

that the denominator of (8) may be written $r_{ht}\sigma_h$ since

$$r_{ht} \equiv r_{(t-j)t} = (\sigma_t - r_{jt}\sigma_j)/\sigma_h \ . \tag{9}$$

Thus (8) may be written

$$r_{j't} = r_{jh}\sigma_h/r_{ht}\sigma_h = r_{jh}/r_{ht} \ . \tag{10}$$

Also, from (6) and (8)

$$r_{j't} = r_{jj}/r_{jt} \ , \tag{11}$$

and finally, from (5) and (8)

$$r_{j't} = r_{tt}r_{jt} \ . \tag{12}$$

REFERENCE

1. Angoff, W. H. Test reliability and effective test length. *Psychometrika*, 1953, **18**, 1-14.

Manuscript received 7/28/55

Revised manuscript received 11/7/55

PSYCHOMETRIKA—VOL. 21, NO. 3
SEPTEMBER, 1956

THE VARIANCE OF THE NUMBER OF MUTUAL CHOICES IN SOCIOMETRY

LEO KATZ
MICHIGAN STATE UNIVERSITY
AND
THURLOW R. WILSON
UNIVERSITY OF NEW MEXICO*

The variance of the number of mutual dyads in a sociometric situation where each member of a group chooses independently and at random is derived for unrestricted numbers of choices per group member, as well as for a fixed number of choices. The distribution of the number of mutuals is considered.

I. *Introduction*

In a sociometric test, a number of the pairs of group members (dyads) may show mutuality. A mutual dyad is defined by the two subjects selecting one another. When each member of the group makes a fixed number of choices, the number of mutual dyads has been assumed to have a binomial distribution with p, the probability of a success, equal to the probability that a given dyad is mutual, and n, the number of binomial observations, equal to the number of dyads (**1**, **7**). Several writers have pointed out that since sampling is without replacement, this is not correct (**2**, **4**, **6**). Let M represent the number of mutual dyads for a group of N numbers. If each member independently makes d selections at random from the $N - 1$ other members, the probability that a given dyad is mutual is $d^2/(N-1)^2$, and the number of dyads is $N(N-1)/2$. The expected value of M is not affected by the non-replacement in sampling and is given by

$$E(M) = np = \frac{Nd^2}{2(N-1)}. \tag{1}$$

We shall show that for a fixed number of choices

$$\text{Var } (M) = E(M)\left(1 - \frac{d}{N-1}\right)^2, \tag{2}$$

rather than the variance appropriate to the binomial distribution,

$$\text{Var (binomial)} = npq = E(M)\left[1 - \left(\frac{d}{N-1}\right)^2\right]. \tag{3}$$

*We are indebted to Robert Bush and Hartley Rodgers of Harvard University for helpful criticisms.

The ratio of the binomial estimate to the variance of (2) is $(N - 1 + d)/(N - 1 - d)$; the binomial formula will appreciably overestimate the variance of M if the size of the group is small.

When the members of the group do not make the same number of choices, it has been assumed that one may approximate the variance of M by using the average number of choices, $\bar{d}$, in the formula for fixed d. We shall develop the expression for the variance of M for the general case of unrestricted numbers of choices and examine the approximation with $\bar{d}$. Finally we consider the distribution of the number of mutual dyads.

II. *Variance with Fixed Number of Choices*

We first derive the variance of M for the case where each member makes the same number of choices, d. Define a random variable X_{ij} for the particular dyad composed of individuals i and j so that X_{ij} is 1 or 0 according as dyad ij is or is not mutual. X_{ij} is a binomial variable; thus

$$\text{Var}\,(X_{ij}) = \frac{d^2}{(N-1)^2}\left(1 - \frac{d^2}{(N-1)^2}\right). \tag{4}$$

Since

$$M = \sum X_{ij}\,, \tag{5}$$

where the sum is taken over all distinct dyads, we have

$$\text{Var}\,(M) = \sum \text{Var}\,(X_{ij}) + 2\sum \text{Cov}\,(X_{ij}X_{ki}). \tag{6}$$

The covariances are summed for all distinct pairs of dyads.

If two dyads are composed of four different individuals, the occurrence of mutual choice for the first dyad is independent of the occurrence of mutual choice for the second; hence the covariance of such a pair is zero. A pair of dyads may not have two members in common, but a pair may have one person in common. There are $N(N-1)(N-2)/2$ such pairs. The expected value of the product $X_{ij}\cdot X_{ik}$ is just the probability that both are equal to 1:

$$E(X_{ij}\cdot X_{ik}) = \frac{d^3(d-1)}{(N-1)^3(N-2)}\,, \tag{7}$$

where dyads ij and ik have one member in common. We find the covariance for an overlapping pair of dyads:

$$\text{Cov}\,(X_{ij}X_{ik}) = -\frac{d^3(N-1-d)}{(N-1)^4(N-2)}. \tag{8}$$

Substituting the values from (4) and (8) in (6) and simplifying we obtain the variance of mutual choices given by (2).

III. *Variance with Unrestricted Numbers of Choices*

We now derive the variance for the case where the members make unrestricted numbers of choices. Let d_i denote the number of choices actually

made by subject i. The probability that dyad ij is mutual is given by

$$p_{ij} = \Pr(X_{ij} = 1) = \frac{d_i\, d_j}{(N-1)^2}. \tag{9}$$

The probability that both dyad ij and kl are mutual is

$$p_{ij,k} = \Pr(X_{ij} \cdot X_{kl} = 1) = \frac{d_i\, d_j\, d_k\, d_l}{(N-1)^4}, \tag{10a}$$

if all four individuals are different, while

$$p_{ij,ik} = \Pr(X_{ij} \cdot X_{ik} = 1) = \frac{d_i(d_i - 1)\, d_j\, d_k}{(N-1)^3(N-2)}, \tag{10b}$$

if the two dyads consist of three different members.

Conventional methods for obtaining the variance (5, pp. 60 ff.) involve computing two sums,

$$S_1 = \sum p_{ij}, \tag{11a}$$

$$S_2 = \sum p_{ij,kl}, \tag{11b}$$

where both sums are evaluated over all distinct sets of subscripts. Considering (5), (9), and (11a) we see that the mean of the number of mutuals is S_1.

To express Var (M) as a function of S_1 and S_2 we first note that

$$\text{Var}(X_{ij}) = p_{ij} - (p_{ij})^2, \tag{12}$$

and

$$\text{Cov}(X_{ij}X_{kl}) = p_{ij,kl} - p_{ij}p_{kl}. \tag{13}$$

Summing over the variances and covariances yields

$$\sum \text{Var}(X_{ij}) = S_1 - \sum (p_{ij})^2, \tag{14}$$

and

$$\sum \text{Cov}(X_{ij}X_{kl}) = S_2 - \sum p_{ij}p_{kl}. \tag{15}$$

Since

$$S_1^2 = \sum (p_{ij})^2 + 2 \sum p_{ij}p_{kl}, \tag{16}$$

substitution of (14) and (15) in (6) produces, after simplification,

$$\text{Var}(M) = S_1 + 2S_2 - S_1^2. \tag{17}$$

We turn next to the computation of S_1 and S_2. The value of S_1 is by definition

$$S_1 = \frac{1}{(N-1)^2} \sum_{i<j} d_i\, d_j. \tag{18}$$

It is, of course, necessary to require $i < j$ in order to prevent duplication of cases in the sum. The summation appearing in (18) is the second elementary sum of the numbers d_i , usually written

$$a_2 = \sum_{i<j} d_i\, d_j \ . \tag{19}$$

These are related to the more familiar power sums,

$$s_m = \sum d_i^m \ , \tag{20}$$

by the relations (3)

$$a_2 = \frac{1}{2}(s_1^2 - s_2), \tag{21a}$$

$$a_3 = \frac{1}{3!}(s_1^3 - 3s_2s_1 + 2s_3), \tag{21b}$$

$$a_4 = \frac{1}{4!}(s_1^4 - 6s_2s_1^2 + 3s_2^2 + 8s_3s_1 - 6s_4). \tag{21c}$$

Those shown will suffice for the present computations. In particular, we have established from (18), (19), (20), and (21a) that the mean number of mutuals is

$$E(M) = \frac{1}{2(N-1)^2}[(\sum d_i)^2 - \sum d_i^2]. \tag{22}$$

If all d_i equal d we obtain (1) above.

The value of S_2 is somewhat more involved. In the first place, each set of four different persons may form mutual pairs in three ways. Secondly, each set of three, distinct, may have any of the three at the center of the chain of two mutual choices. Taking both of these features into account, we have

$$\begin{aligned} S_2 = {} & \frac{3}{(N-1)^4} \sum_{i<j<k<l} d_i\, d_j\, d_k\, d_l \\ & + \frac{1}{(N-1)^3(N-2)} \sum_i \{ \sum_{\substack{j<k \\ \neq i}} d_i(d_i - 1)\, d_j\, d_k \}. \end{aligned} \tag{23}$$

Making use of (21a) and (21b) we have, after some reduction,

$$\begin{aligned} S_2 = {} & \frac{1}{8(N-1)^4}(s_1^4 - 6s_1^2s_2 + 3s_2^2 + 8s_1s_3 - 6s_4) \\ & + \frac{1}{2(N-1)^3(N-2)}(s_1^2s_2 - s_2^2 - 2s_1s_3 + 2s_4 - s_1^3 + 3s_1s_2 - 2s_3). \end{aligned} \tag{24}$$

Combining this with previous computations, we obtain

$$\begin{aligned}\text{Var}\,(M) &= \frac{1}{2(N-1)^4}(-2s_1^2s_2 + s_2^2 + 4s_1s_3 - 3s_4) \\ &+ \frac{1}{(N-1)^3(N-2)}(s_1^2s_2 - s_2^2 - 2s_1s_3 + 2s_4 - s_1^3 + 3s_1s_2 - 2s_3) \qquad (25) \\ &+ \frac{1}{2(N-1)^2}(s_1^2 - s_2).\end{aligned}$$

When all d_i equal d, (25) reduces to (2).

As examples with unequal numbers of choices, consider two sociometric measurements on 10 individuals. In the first case, two subjects choose two persons each, five choose three, and three choose four. The power sums are 31, 101, 343, and 1205 for s_1, s_2, s_3, and s_4, respectively. Equations (22) and (25) give $E(M)$ equal to 5.31 and Var (M) equal to 2.30. Observe that the use of $\bar{d}$ equal to 3.1 in (1) and (2) would give the very close approximations, $E(M)$ equal to 5.34 and Var (M) equal to 2.29. This happy situation would not obtain if the numbers of choices were considerably more variable, as in the second example; five make a single choice and five make seven choices. The power sums are 40, 250, 1720, and 12,010. From (22) and (25) $E(M)$ is 8.33 and Var (M) is 2.33. With $\bar{d}$ of 4, we obtain the values 8.89 for $E(M)$ and 2.74 for Var (M).

IV. *Distribution of the Number of Mutual Dyads.*

In principle, we could determine the distribution of M exactly by conventional methods (**5**, p. 64). Denote the maximum possible number of mutual dyads for a group by Max. We define S_m as suggested by (11a) and (11b) above. Then

$$\begin{aligned}\Pr\,(M = m) = S_m - \binom{m+1}{m}S_{m+1} + \binom{m+2}{m}S_{m+2} \\ - \cdots - + \binom{\text{Max}}{m}S_{\text{Max}}\,.\end{aligned} \qquad (26)$$

We have used (26) only for the case where each person makes a single choice, where

$$S_m = \frac{\binom{N}{2m}\binom{2m}{2}\binom{2m-2}{2}\cdots\binom{2}{2}}{m!(N-1)^{2m}}, \qquad (27)$$

and Max equals $N/2$.

We conjecture that for large groups with roughly equal d, M has an Approximately Poisson distribution; from (2) one readily notes that with

increasing group size and r held constant, the variance of M approaches $E(M)$ and the covariance term approaches zero.

The senior author is currently engaged in determining higher moments for various common fixed d by methods similar to those used to produce the variance for the general case. These higher moments will permit determination of the distribution of M for small groups.

REFERENCES

1. Bronfenbrenner, U. A constant frame of reference for sociometric research. *Sociometry*, 1943, **6**, 363–397.
2. Crisswell, Joan H. Foundations of sociometric measurement. *Sociometry*, 1946, **9**, 7–13.
3. David, F. N. and Kendall, M. G. Tables of symmetric functions—part I. *Biometrika*, 1949, **36**, 431–449.
4. Edwards, Daisy S. The constant frame of reference problem in sociometry. *Sociometry*, 1948, **11**, 372–379.
5. Feller, W. Probability theory and its applications. Vol. 1, New York: Wiley, 1950.
6. Loomis, C. P. and Pepinsky, H. B. Sociometry, 1937–1947: theory and methods. *Sociometry*, 1948, **11**, 262–286.
7. Moreno, J. L. and Jennings, Helen H. Sociometric measurement of social configurations, based on deviations from chance. *Sociometry Monographs*, No. 3, 1945, Beacon House.

PSYCHOMETRIKA—VOL. 21, NO. 3
SEPTEMBER, 1956

A NOTE ON JENKINS' "IMPROVED METHOD FOR TETRACHORIC *r*"

JOSHUA A. FISHMAN

COLLEGE ENTRANCE EXAMINATION BOARD

Some readers who will be delighted to utilize Jenkins' method and tables for estimating the tetrachoric correlation (1) may be puzzled to discover that no explicit provision for negative correlations is included. If we follow Jenkins' instruction to "letter the fourfold table so that a is smaller than d and ad greater than bc" four possible arrangements may obtain:

$$1.\ \begin{array}{c|c} c & d \\ \hline a & b \end{array} \qquad 2.\ \begin{array}{c|c} b & a \\ \hline d & c \end{array} \qquad 3.\ \begin{array}{c|c} a & b \\ \hline c & d \end{array} \qquad 4.\ \begin{array}{c|c} d & c \\ \hline b & a \end{array}$$

Of the above four arrangements, the first (which is the one illustrated by Jenkins) and the second are indicative of a positive correlation and the third and fourth are indicative of a negative correlation. If either of the latter two arrangements does obtain, then the final correction [obtained by multiplying the base correction by the multiplier, as in step 5, p. 257, (1)] should be (algebraically) *added to*, rather than subtracted from the negative uncorrected r to obtain the corrected tetrachoric r, which should, of course, be given a negative sign. The important fact to keep in mind is that the correction always reduces the *absolute* size of the uncorrected tetrachoric r.

A few words also might be in order concerning the location of decimal points in Tables 2 (Base Correction) and 3 (Multipliers for Base Correction). Whereas in the former table the omitted decimal points consistently belong before the first digit of the reported three-digit table entries, in the latter table the omitted decimal points belong before the first digit of two-digit table entries, between the first and second digit of three-digit table entries, and followed by a zero for one-digit table entries. Thus, whereas a table entry of 106 should be understood as .106 in Table 2, it should be taken as 1.06 in Table 3. Furthermore, 90 is .90 (as illustrated by Jenkins, p. 257) but 9 is .09 in Table 3.

1. Jenkins, W. L. An improved method for tetachoric *r*. *Psychometrika*, 1955, 20, 253-258.

BOOK REVIEW

Mathematical Models of Human Behavior. Proceedings of a Symposium Sponsored by Dunlap and Associates, Inc., and the Commission on Accidental Trauma, Armed Forces Epidemiological Board. 1955. vii + 103 pp.

Those who expect to read this book in its entirety, or nearly so, would do well to turn first to Professor Lazarsfeld's Concluding Remarks, since they provide some degree of unification for what is otherwise a rather disjointed collection of reports on several diverse lines of investigation. Those who do not expect to read the entire book would do well to read at least Professor Lazarsfeld's Concluding Remarks for a brief but lucid statement of the function of models, especially in the behavioral sciences. He distinguishes between static and dynamic models, makes passing reference to the predictive function of models, and concentrates on their linguistic function. The linguistic function he divides into three parts: organizing, analytical, and mediating. Naturally his remarks refer to the papers in the Symposium, but they are quite intelligible in themselves.

The Symposium itself was held in February of 1954, in connection with a study being made by Dunlap and Associates, Inc., on the application of mathematical techniques to the study of accidents. Since accidents are "partly the result of human behavior," and since many experts were already engaged in devising and studying mathematical models of human behavior, the proposal was made to invite some of these experts to meet together to describe their work and to participate in informal discussions. This publication contains the papers, but unfortunately not the discussion.

Of the ten papers included (not counting the Concluding Remarks, and not counting a paper by Lorge and Solomon to be published elsewhere) only one deals with accidents. This is one by H. H. Jacobs, who discusses the difficulties in trying to separate the effects of contagion from individual differences as to liability. Two of the speakers, Bush and Estes, discussed somewhat related stochastic learning models. The other seven papers had to do, more or less directly, with utility, or decision making, or both.

For general background on models, the paper by Coombs and Kao might be classed along with Lazarsfeld's Concluding Remarks. This is actually the first section of a report on multidimensional analysis, and one feels suspended in mid-air at the end. As an inducement to learn more, this paper succeeds very well. At the other extreme, a paper by Luce on the formation of coalitions in game theory is largely for the experts. The other five papers are more self-contained, and more directly concerned with utility and decision as such.

Professor Lazarsfeld gently chides the speakers for their preoccupation with gambling, although, as one of them remarks, the gambling situation provides the most direct and realistic contact with the individual's utility function. Merrill Flood's little "Group Preference Experiment" does not make use of a conventional gambling situation, but it does afford, to each of the subjects, a possibility, without certainty, of some gain. A collection of objects is shown to a group of individuals, and certain broad conditions are laid down according to which one of these objects can be had for the group to dispose of. They are left to decide which object it shall be and how it shall be disposed of. Disposition might be by lot to one of the group, or by auction or sale with proceeds divided, and other possibilities can be conceived.

Marschak is concerned with decisions by individuals, pointing out that even when the outcome of an act is known with certainty the same individual may make different choices at different times. He considers various hypotheses that might be made in formulating a model of such inconsistent behavior. Markowitz discusses a hypothesis of Friedman and

Savage for explaining insurance and lotteries; Markowitz replaces it by one of his own which seems to accord better with well-known facts. Jarvik's discussion of gambling is largely discursive, and Edwards describes a series of experiments on gambling which endeavor to arrive at utilities.

The success of the Symposium as such could be judged best from the discussion which is not published. As a publication, it is interesting for showing the diversity of activities under way, but unsatisfying just because of the diversity. As an issue of a periodical, this little volume would do very well. It lacks the cohesion to stand well by itself.

The proofreading is rather poor; presumably the spelling "baracentric" on page 22 is a typographical error.

A. S. HOUSEHOLDER